Model-based Analysis of Translucent Objects using Spatially and Temporally Modulated Light

Submitted to
Graduate School of Information Science and Technology
Osaka University

January 2017

Kenichiro TANAKA

Abstract

Measurement of transparent or translucent objects is an important technology with broad potential applications. In the computer vision field, computational photography approaches that combine optical design and computational algorithms to obtain informative images have been actively developed to enhance and restore images. However, it remains difficult to recover invisible information contained within transparent or translucent objects because light penetrates and scatters inside the object, heavily degrading observed images.

Our goal is to recover invisible information that has been degraded by scattering and is too complex to be directly measured and analyzed. In this thesis, we propose a measurement framework using either spatially or temporally modulated light that conveys the frequency responses of the translucent objects. We model the spread of modulated light at several frequencies. Based on this model, we recover invisible information from translucent objects by processing the frequency responses.

Using a projector to realize spatial modulation, we reveal how the pattern projection conveys the information of spatial spread of light. Using a Time-of-Flight (ToF) camera to realize temporal modulation, we propose a model for the distortions caused by the temporal spread of light. Based on these relationships, we recover clear images stored inside translucent objects, accurate 3-d shape, and material. We demonstrate the qualitative and quantitative effectiveness of our method in real-world experiments. Herein, we show that both spatial and temporal spread of light can be analyzed using modulated light.

List of Publications

A. Award

1. Honorable Mention, IEEE International Conference on Computational Photography, 20th Apr., 2013.

B. Journal Paper

1. **Kenichiro Tanaka**, Yasuhiro Mukaigawa, Hiroyuki Kubo, Yasuyuki Matsushita, Yasushi Yagi, "Recovering Inner Slices of Layered Translucent Objects by Multi-frequency Illumination", IEEE Transaction of Pattern Analysis and Machine Intelligence (TPAMI), *Accepted*.
2. 田中 賢一郎，向川 康博，八木 康史，"平行高周波照明による透視画像の散乱光除去"，電子情報通信学会論文誌 D，Vol. J96-D，No. 8，pp.1834-1843，2013.

C. International Conference (Refereed)

1. **Kenichiro Tanaka**, Yasuhiro Mukaigawa, Hiroyuki Kubo, Yasuyuki Matsushita, Yasushi Yagi, "Recovering Transparent Shape from Time-of-Flight Distortion", Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4387–4395, Jun. 2016.
2. **Kenichiro Tanaka**, Yasuhiro Mukaigawa, Hiroyuki Kubo, Yasuyuki Matsushita, Yasushi Yagi, "Recovering Inner Slices of Translucent Objects by Multi-frequency Illumination", Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5464–5472, Jun. 2015.

3. **Kenichiro Tanaka**, Yasuhiro Mukaigawa, Yasuyuki Matsushita, Yasushi Yagi, “Descattering Transmissive Observation using Parallel High-frequency Illumination”, Proceedings of the IEEE International Conference on Computational Photography (ICCP), pp.96–103, Apr. 2013.

D. International Workshop

1. **Kenichiro Tanaka**, Yasuhiro Mukaigawa, Yasuyuki Matsushita, Yasushi Yagi, “Recovering Transparent Shape from Time-of-Flight Distortion”, The 11th International Workshop on Robust Computer Vision, (IWRCV), Total：1 page, Nov. 2016.
2. **Kenichiro Tanaka**, Yasuhiro Mukaigawa, Yasuyuki Matsushita, Yasushi Yagi, “Descattering Transmissive Images of Living Body using Parallel High Frequency Illumination”, The 8th International Workshop on Robust Computer Vision, (IWRCV), Total： 1 page, Jan. 2014.
3. **Kenichiro Tanaka**, Yasuhiro Mukaigawa, Yasuyuki Matsushita, Yasushi Yagi, “NIR Parallel High-frequency Illumination for Descattering”, Project meeting of G3 alliance, Total： 1 page, Oct. 2013.
4. **Kenichiro Tanaka**, Yasuhiro Mukaigawa, Yasuyuki Matsushita, Yasushi Yagi, “Descattering of Transmissive Image using Parallel High-frequency Illumination”, A Joint Workshop of Osaka-Univ. and Peking-Univ. Groups, Total： 1 page, Mar. 2013.
5. **Kenichiro Tanaka**, Yasuhiro Mukaigawa, Yasuyuki Matsushita, Yasushi Yagi, “Sharpening Transmissive Images using Parallel High-frequency Illumination”, The 7th International Workshop on Robust Computer Vision (IWRCV2013), Total: 1 page, Jan., 2013.

E. Domestic Conference (Refereed)

1. 田中 賢一郎，向川 康博，八木 康史，“平行高周波照明による透視画像の鮮明化”，第 15 回 画像の認識・理解シンポジウム (MIRU2012)，OS10-2，ページ数: 8 ページ， 2012 年 8 月.

F. Domestic Workshop

1. **Kenichiro Tanaka**, Yasuhiro Mukaigawa, Hiroyuki Kubo, Yasuyuki Matsushita, Yasushi Yagi, “Recovering Transparent Shape from Time-of-Flight Distortion” (CVPR2016), 画像の認識･理解シンポジウム (MIRU2016) 招待講演 , ページ数: 0 ページ, Aug. 2016.
2. **Kenichiro Tanaka**, Yasuhiro Mukaigawa, Hiroyuki Kubo, Yasuyuki Matsushita, Yasushi Yagi, “Recovering Inner Slices of Translucent Objects by Multi-Frequency Illumination” (CVPR2015), 画像の認識・理解シンポジウム (MIRU2015) 招待講演 , ページ数: 0 ページ, Jul. 2015.
3. 田中 賢一郎，向川 康博，久保 尋之，松下 康之，八木 康史，“多重高周波照明による半透明物体内部のスライス可視化”，情処研報 CVIM 195，ページ数: 8 ページ，2015 年 1 月.
4. **Kenichiro Tanaka**, Yasuhiro Mukaigawa, Yasuyuki Matsushita, Yasushi Yagi, “Recovering Slices inside Translucent Media using Coaxial Multi-frequency Illumination”, 第 17 回 画像の認識・理解シンポジウム (MIRU2014), ページ数: 2 ページ, Aug. 2014.
5. 田中 賢一郎，向川 康博，八木 康史，“プロジェクタ・カメラシステムによる散乱体内部の鮮明化”，附置研究所間アライアンスによるナノとマクロをつなぐ物質・デバイス・システム創製戦略プロジェクト成果報告会，ページ数: 1 ページ，2014 年 5 月.
6. 田中 賢一郎，向川 康博，松下 康之，八木 康史，“平行高周波照明による透過光と散乱光の分離”, FIRST シンポジウム「科学技術が拓く 2030 年」へのシナリオ, ページ数: 1 ページ, 2014 年 2 月.
7. 田中 賢一郎，向川 康博，松下 康之，八木 康史，“平行高周波照明による透過光と散乱光の分離”, 第 69 回学術講演会, ページ数: 1 ページ, 2013 年 12 月.
8. 田中 賢一郎，向川 康博，松下 康之，八木 康史，“平行高周波照明における周波数の適応的選択”, 第 16 回 画像の認識・理解シンポジウム (MIRU2013), ページ数: 2 ページ, 2013 年 8 月.
9. 田中 賢一郎，向川 康博，松下 康之，八木 康史，“透視画像の鮮明化のための平行高周波照明システムの開発”，附置研究所間アライアンスによるナノとマクロをつなぐ物質・デバイス・システム創製戦略プロジェクト成果報告会，ページ数: 1 ページ，2013 年 4 月.
10. 田中 賢一郎，向川 康博，八木康史，“平行高周波照明による透過光の抽出”，情処研報 CVIM 182，ページ数: 8 ページ, 2012 年 5 月.

Acknowledgements

First and foremost, I am much obliged to my adviser Professor Yasushi Yagi for his support and advice throughout my research at Osaka University. I appreciate his comments for my research vision, way of thinking, method, and career plan. The laboratory has a comfortable research environment and excellent members, and I have been really lucky to be a member of the laboratory.

I am forever indebted to Professor Yasuhiro Mukaigawa for his guidance and unwavering support throughout my time at the graduate school. He has given me a lot of critical, important, meaningful advice. This work would have never been possible without him. I also appreciate that he accepted me as a special research student at his laboratory at the Nara Institute of Science and Technology.

Professor Yasuyuki Matsushita has been an invaluable and respected collaborator for me. His straightforward and intelligent approach and world-class skill have strongly influenced me. I am incredibly grateful to him for his constructive discussions and contributions to our collaborative work, especially in the mathematical, writing, and presentation aspects.

I thank all the researchers working at Professor Yagi's laboratory at Osaka University, especially Professor Yasushi Makihara, Professor Daigo Muramatsu, Professor Ikuhisa Mitsugami, and Professor Fumio Okura. I really appreciate their instruction, comments, and assistance. I thank other collaborators including Professor Hajime Nagahara at Kyushu University and Mr. Hajime Mihara for their excellent work. I also thank Professor Shinji Kusumoto for being on my dissertation committee.

I appreciate all supports from secretaries working in Professor Yagi's laboratory including accounting, office work, and management of my grant. I thank Ms. Noriko Yasui, Ms. Masako Kamura, Ms. Makiko Fujimoto, Ms. Masako Sugimoto, Ms. Kumiko Nakagawa. I also thank Ms. Toshie Nobori, a secretary with Professor Mukaigawa's laboratory, for miscellaneous businesses.

I wish to express my appreciation to the member of photometric analysis group for their valuable time and assistance. I thank Dr. Seiichi Tagawa, Dr. Chika Inoshita, Dr. Md. Abdul

Mannan, Mr. Tsuyoshi Takatani, Mr. Shosei Moriguchi, Mr. Ryo Matsumoto, Mr. Takahiro Matsumura, Mr. Sho Ikemoto, Mr. Takanori Okamoto, Mr. Yuya Iwaguchi, and Mr. Kazuya Kitano. I also offer my appreciation to laboratory members and alumni, including Dr. Haruyuki Iwama, Dr. Mitsuru Nakazawa, Mr. Kohei Shiraga, and Mr. Ryo Kawai.

I also want to acknowledge the many students, postdocs, and professors of other laboratories that have enriched my life. I thank Professor Takuya Funatomi for fruitful comments on my research, and for fishing. I thank Professor Hiroyuki Kubo for his work and knowledge of graphics and rendering. Special thanks to Dr. Takahito Aoto for the life hacks, delicious dinner, ramen noodles, udon noodle tour, skiing, public bath visits, and other activities. I thank Professor Yuta Nakashima, Professor Norihiko Kawai, Mr. Hikari Takehara, Ms. Mayu Otani, Mr. Atsushi Takabe, Mr. Ryoichiro Nishi, and the other members of Professor Yokoya's laboratory at NAIST for their kindness, and for dinner. I also thank Professor Mukaigawa's laboratory members, including Mr. Takafumi Iwaguchi, Mr. Shigenobu Asada, Ms. Aoi Ishihara, Mr. Takahiro Kushida, Mr. Fusataka Kuniyoshi, Mr. Shun Sakurai, Ms. Karin Tsuda, and Ms. Yuki Sato, for their work and for their leisure.

A part of this work is supported by JSPS KAKENHI Grant-in-Aid for JSPS Fellows, Grant Number JP26-6433, from the Japan Society for the Promotion of Science, Japan.

最後に，私の進路を理解し，忍耐強く応援していただき，また経済的にサポートしてくれた両親，泰宏，真理子に感謝します．ありがとう．

Contents

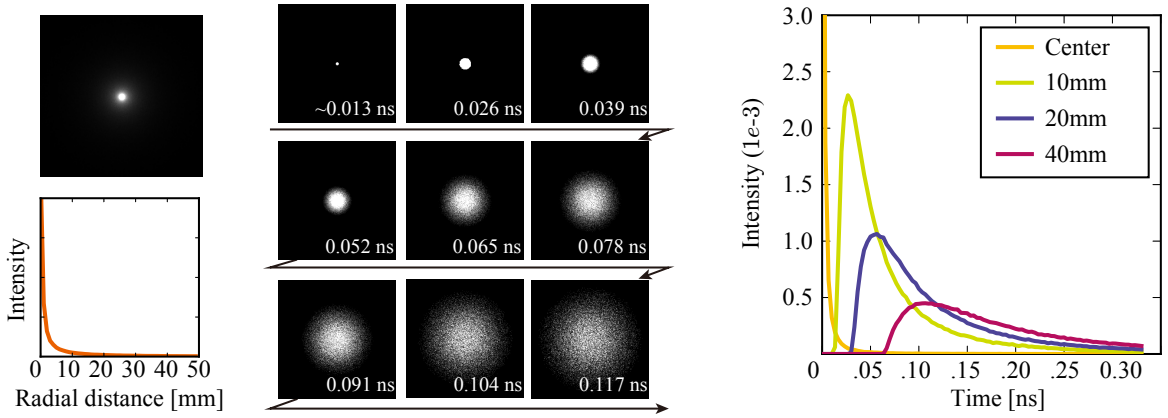
Chapter 1

Preface

Measurement of transparent and translucent objects is an important technology that may have applications in fields, including autonomous robotics, industrial inspection, underwater science, art, and historical research. However, such measurements remain challenging in computer vision applications because of the complicated optical phenomena involved, such as subsurface scattering and refraction. For example, when a translucent object is illuminated by a laser pointer, the laser beam becomes blurred and degraded as a result of the light penetrating and scattering inside the object. To overcome these problems in measuring transparent and translucent objects, their complicated optical responses, referred to as 'light transport', must be compensated for.

Properly dealing with the light transport of a scene can help a computer to correctly understand the scene. For example, clear images can be recovered in foggy weather if the scattering properties of the fog are identified. Translucency of the scene is also important for recognizing the scene, as well as removing its effects. Taking into account the magnitude of subsurface scattering, the material of the target object can be recognized because translucency conveys the optical properties of the material.

The spread of light due to scattering, which is part of light transport, is often analyzed using the point spread function (PSF) as shown in Fig. 1.1(a). This represents the response of the infinitesimal impulse light. Spatial PSF represents the spatial spread of the light; hence, contributions of direct rays can be recovered via deconvolution [1]. In a similar manner, temporal PSF as shown in Fig. 1.1(c) represents the spread of light with time, where scattering rays distribute temporally according to their travelled path length. In addition, temporal PSF provides information about direct paths because the speed of light varies according to the refractive index, whereas spatial PSF does not provide such information. Figure 1.1 shows simulated PSFs rendered by Monte Carlo ray tracing as an example of the spread of light inside diluted milk. Light is spread spatially and temporary, carrying rich information on the target object.



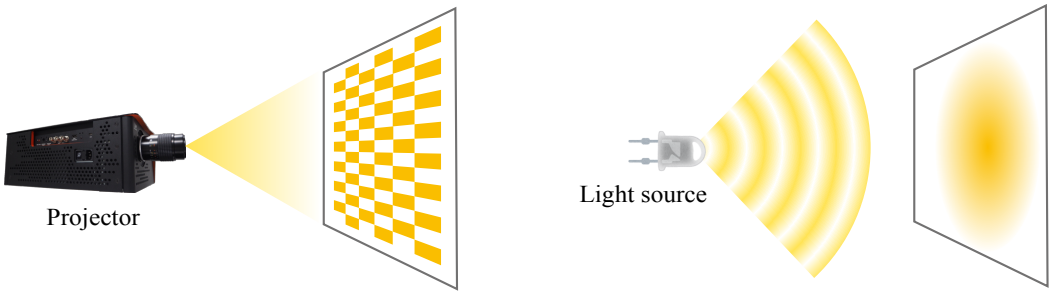
(a) Spatial PSF (b) Spatiotemporal PSF (c) Temporal PSF

Figure 1.1: Point spread functions (PSFs) of diluted milk rendered by Monte Carlo ray tracing. Parameters, including coefficients of scattering and absorption, are chosen from [5], and we assume the target is an infinite plane of infinite thickness. (a) An ordinary (spatial) PSF. This represents the spatial spread of light in a steady state. (b) Spatiotemporal PSFs. They represent the transition of the spread of light spatially and temporally. (c) A temporal PSF at points 0 mm, 10 mm, 20 mm, and 40 mm from the center. Temporal PSF becomes broader at larger radial distances. Both spatial and temporal spreads of light convey physical information of the target. Analyzing these light behaviors will make the scene more understandable.

Some methods directly measure the spread of light, such as measuring spatial PSF [2], BSSRDF [3], and temporal PSF as known as light-in-flight [4]. Using measured PSFs, the original information can be recovered from observations blurred by scattering, using deconvolution, inverse rendering, and optimization. Directly measuring the light transport, however, suffers from the trade-off among measuring time, cost, and data quality, because light transport is high dimensional. Light transport is a combination of the position, angle, timing, wavelength, and coherence of both the incident light and the output light. Therefore, approaches for direct measurement of the spread of light are insufficient for realistic applications. Improved techniques to analyze light transport are required.

Our Approach

We aim to recover invisible information from images degraded by complicated light transport such as scattering. Because scattering is too complex to be fully measured and analyzed, we do not explicitly recover either the scattering parameters or the PSFs. Instead, we propose model-based approaches for visualizing invisible information that indirectly observes spatial



(a) Spatial modulation (b) Temporal modulation

Figure 1.2: Amplitude-modulated illumination. (a) High-frequency illumination, in which the illumination is spatially modulated by a projector. Illumination is rapidly switched on and off with the illuminated position. A typical wavelength is sub-millimeter. (b) Illumination is temporally modulated at several tens of mega-hertz, using a ToF camera.

and temporal PSFs. For some specific applications, full recovery of PSF is redundant; thus, the measurement cost can be reduced by modeling the PSFs and observation process.

We adopt either spatially or temporally amplitude-modulated illumination as shown in Fig. 1.2. Spatially modulated illumination can be realized using a projector, which can control the amplitude of light by projecting a high-frequency pattern. Temporal modulation can be realized by turning the light source on and off at high speed, which is achieved using a Time-of-Flight(ToF) camera. Observation under an ordinary illumination such as a light bulb degrades the information of the scene because scattering behaves as a low-pass filter. In contrast, observation under amplitude-modulated light maintains the frequency response of the scene. Using various frequencies of the amplitude-modulated illumination, many frequency responses can be obtained; thus, invisible information can be recovered via computation.

We propose a framework that simultaneously deals with spatially and temporally modulated illumination at different frequencies, as shown in Fig. 1.3. We illuminate the scene by spatially or temporally amplitude-modulated light and observe the amplitude and phase shift of the returned wave. We repeat observations at various illumination frequencies. From these observations, we analyze the scene to recover invisible information based on the model about the PSFs and modulated illumination.

Implementations and Contributions

As an implementation of our framework, we propose spatial and temporal versions of our concept. A high-frequency spatial pattern is projected onto the scene and shifted progressively,

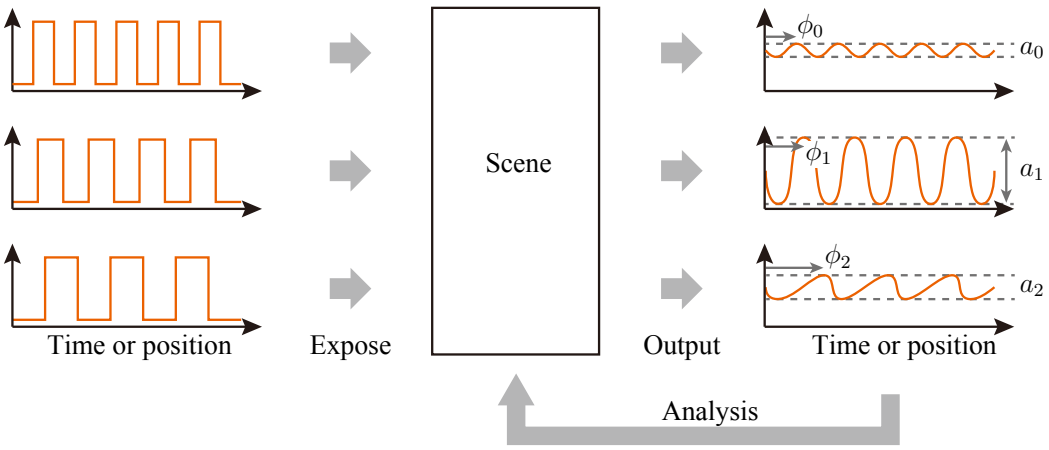


Figure 1.3: An overview of our framework. We illuminate the scene using either spatially or temporally amplitude-modulated light at different frequencies and observe the response as amplitude and phase shift of the returned wave. From these observations, we recover invisible information from the scene via computation based on the distortion model of the returned wave.

and the amplitude of the remaining pattern is observed. We show the importance of the optical design for high-frequency illumination and recover invisible information from translucent objects based on the relationship between the spatial PSFs and the multi-frequency illumination. Furthermore, we use a ToF camera for temporally modulated observations and model the measurement distortion due to temporal PSFs. Using the distortion model, we recover the material and shape of visually similar objects.

The main contribution is that we recover both spatial and temporal cues for scene analysis without explicitly recovering PSFs of the target objects, using to our model-based approach. This approach uses a combination of the optical system design and computational algorithms, and it does not require any expensive direct measurement of PSFs. We show that the scene, including translucent objects, can be analyzed using our model of the relationship between PSFs and frequency responses to spatially and temporally modulated light.

Organization of thesis

The thesis is organized as follows. In Chapter 2, we summarize the related work. In Chapter 3, we present the theory of optical design and high-frequency illumination from the viewpoint of pixel correspondence between the projector and the camera. Results of sharpening images of transmissive observations are shown. In Chapter 4, we reveal the relationship between the spatial PSFs and spatially modulated light at different frequencies. We recover hidden layers

inside translucent objects, where spatial PSF increases with increasing depth. In Chapter 5, we analyze the temporal version of our concept. We show that a ToF camera, which uses temporally amplitude-modulated light, observes the distorted phase according to the refraction. We recover the shape of a transparent object using the distorted phase of the returned wave. In Chapter 6, we propose a model for temporally amplitude-modulated light and temporal PSFs at various frequencies. We classify the material of the scene based on this proposed model. We conclude this thesis in Chapter 7. including limitations and future work.

Chapter 2

Related Work

In this chapter, we summarize the related work. Firstly, we review the direct measurements of the slice of global light transport. Secondly, we summarize active illumination works that illuminate the scene by amplitude-modulated light. They are strongly related to our work because our aim is to analyze spatial and temporal light transport using amplitude-modulated light. Related to our specific implementations, we summarize the work of descattering, separating and rendering layered translucent objects, shape recovery of transparent and translucent objects, and material classification, respectively.

Direct Measurement of Light Transport

The light transport is a high dimensional effect because it is a combination of light's position, angle, wavelength, timing, and coherence for both incident and output lights. Therefore, direct measurement of light transport is generally upper bounded by trade-offs among complexity of measurement system, cost, resolution of data, and time. A slice of the light transport along with the spatial spread is represented as point spread functions and measured by infinitesimal pointed illumination [2]. Angular slice of the light transport is called as Bidirectional Reflectance Distribution Function (BRDF). A naïve direct measurement of BRDF is using a gonioreflectometer [6] and multi-camera system [7], which directly measures the angular distribution of reflection. BRDF can be also measured using ellipsoidal mirror [8], using a sphere [9], and dipole model fitting [10]. Measureing BRDF is generally upper bounded by the trade-off of measuring time and resolution. Adding spatial information to the BRDF is called Bidirectional Scattering Surface Reflectance Distribution Function (BSSRDF). 8-D BSSRDF or 8-D reflectance field is also directly measured using special mirror devices [3, 11], which earns spatial resolution by sacrificing angular resolution. A temporal slice of the light transport is called as light-in-flight and transient imaging, which visualizes the temporal propagation of

light of the scene. The light-in-flight can be measured directly using a femtosecond-pulsed laser and a streak camera [4], which are enormously expensive devices. We avoid the direct measurement of light transport and use relationships between the light transport and active illumination for analyzing light transport.

Spatially Coded Illumination

Our method can be grouped in a class of active sensing techniques that use high-frequency pattern projection. These techniques are based on the relationship between the pattern projection and the spatial spread of light. The original Nayar *et al*. [12]'s method separate direct and global components by projecting multiple high-frequency patterns, where direct components include specular and diffuse reflection and global components include inter-reflection, volume and subsurface scattering. Extending this method, Lamond *et al*. [13] separate specular and difuse refrection, Mulaigawa *et al*. [14, 15] separate single and multiple scattering, Tanaka *et al*. [16] separate transmissive and scatttered lights, and Tanaka *et al*. [17] separate upper and inner layers. We summarize these method in terms of what is exactly the direct components in the viewpoint of optical design in this thesis. In addition, Achar *et al*. [18, 19] overcome the problem of motion and defocus of high-frequency illumination by total variation optimization and changing projector's focus while projection. Reddy *et al*. [20] separate light transport into direct, near-indirect, and far-indirect rays by frequency-domain modeling and analysis. There are other pattern projection techniques to decompose light transport. Gupta *et al*. [21] acquire scene depths with direct-global separation by modeling both projector's defocus and global light transport. O'Toole *et al*. [22, 23, 24] illuminate a scene by a pattern while masking the camera by the complementary pattern to spatially probing the light transport of the target scene. Our methods explained in Chapters 3 and 4 are also based upon Nayar's method [12] and we use a relationship among direct components, the size of scattering blur, and the pitch of the projection pattern to separate depth layers. Especially in Chapter 4, we use the pattern projection not only for separating scattering effects, but also for recovering appearance of multiple inner slices.

Our work is also related to imaging through scattering/occlusion methods in computational photography. Levoy *et al*. [25] combine a synthetic aperture technique [26] in remote sensing and a confocal imaging technique [27] in microscopy to reconstruct images of scenes behind occlusion or inside scattering medium. Fuchs *et al*. [28] also use the confocal imaging for the purpose of recording a solid target in scattering media. Because these methods are based on

confocal imaging, for which both the camera and projector need exact focusing on the same depth plane, it is needed to have a priori knowledge about the target depth. For imaging through scattering media, Narasimhan *et al*. [29]'s and Gu *et al*. [30]'s methods sharpen images of a target scene in muddy liquid by precisely modeling single scattering. Their methods work well for those scenes that do not exhibit multiple scattering. Differently from these works, our method recovers images inside an optically thick translucent object, where significant multiple scattering is observed and the optical thickness of the target is unknown.

Temporally Coded Illumination

A time domain impulse response of the scene, as known as light-in-flight and transient imaging, can be obtained using an interferometer [31], holography [32, 33], and femtosecond-pulsed laser [34, 4, 35]. The time domain impulse response can be also recovered using the ToF camera. Because the ToF camera is a device for measuring sub-nano second phenomena, it can be used for visualizing the light propagation of the scene by frequency sweep [36, 37, 38] and optical coding [39, 40], while it requires customization of a ToF camera. These measurement methods may be able to be applied to the task of material classification, although they require careful and expensive setups. On the other hand, our method bypasses the exact recovery of the time domain impulse response and simply uses the measured depth of a ToF camera.

When a ToF camera measures a multi-path scene, the measured depth is distorted due to inter-reflections and subsurface scattering, known as the multi-path interference. Mitigating the multi-path interference and recovering the correct depth is of broad interest, and it has been studied by assuming two-bounce and simplified reflection model [41, 42, 43, 44], parametric model [45, 46], K-sparity and optimization [47, 48, 49], stereo ToF cameras [50], using external projector [51], and frequency sweep [52]. Instead of recovering the correct depth, we use a distorted depth as a cue for the material classification in Chapter 6. We show that, once the material classification has been achieved, the classification result can be used for correcting depths. Besides, we use a distorted depth due to the refractive index of the medium as a cue for the transparent shape recovoery in Chapter 5.

There are other scene analysis methods using ToF cameras, *e.g.*, recovering the shape of transparent and translucent objects [53, 54], and measuring a slice of BRDF [55]. In addition, computational imaging methods using the ToF camera, such as imaging around the corner [56, 57], separating direct and indirect light transport [58, 40, 59], imaging the velocity of the object [60, 61], and imaging at a specific depth [62] are proposed. Our method can also be considered one of the scene analysis methods as it aims at material classification of the scene.

Descattering

X-ray imaging is used for obtaining clear transmissive images, because of its high-transmissive and low-scattering characteristics. However, it is known that X-ray have a risk of radiation exposure, and alternative imaging techniques are wanted. As a safe imaging technique, near infra-red light is used for living body imaging because of its high transmittance. Vein authentication for security purpose is one example of this type of imaging. In the medical field, the difference in the absorption properties of near infrared light between oxy-hemoglobin and deoxy-hemoglobin are used to measure blood oxygen levels [63]. Our goal is to obtain sharp images by removing scattering lights by using visible or infrared light and special optics.

In the optical measurement field, parallel light is traditionally used. Shadowgraphy [64] is a volumetric density distribution imaging method based on the detection of slight directional change of light rays caused by differences of refractive index in the medium. Although schlieren method [64] is similar to shadowgraphy, it cuts some parallel rays at their focal point with a knife edge and can visualize the density distributions with higher contrast. Using this optical design, the schlieren method has been used for measuring scattering property of the medium [65]. It is introduced to the computer vision field ,*e.g.*, schlieren photography with light field probes [66] has been proposed to produce ray-to-ray correspondences. In Chapter 3, we also use the observation that a parallel ray changes its direction because of scattering, and the optical setting is also similar. However, while shadowgraphy and schlieren methods use only the directional clue, we use both the directional and positional clues by combining them with high-frequency illumination. Therefore, we can separate the transmissive and scattered lights more effectively.

There are various approaches for descattering in the field of computational photography via software post-processing and/or specially designed imaging setups. Gilvert and Pernicka [67] and Treibitz and Schechner [68] show that circular polarizers can be used for reducing back-scattering from the scene. The techniques are shown to be useful; however, the methods are limited to scenes where scattered light rays are well polarized. Narasimhan *et al*. [29]'s method sharpens images of a scene in muddy liquid by precisely modeling of single scattering, which works well for the scenes that do not contain multiple scattering. Fuchs *et al*. [28] use a projector for descattering. Kim *et al*. [69] propose a method for sharpening transmissive images using angular information obtained by using a light field camera. Wu *et al*. [58] use a femto-laser and a streak camera for image decomposition to achieve a similar goal. Our descattering method is based on the high-frequency illumination combining with parallel optics and polarizers more efficiently, which uses positional, angular, and polarization cues about the spread of light.

Dehazing from a single image is also actively studied, which improve the contrast of the image and ignore scattering blur. It is solved by assuming some prior knowledge including dark-channel prior [70], different statistics between color and depth [71], local color statistics [72], and non-local color statistics [73]. They are however difficult to recover the correct color of the scene because the problem is highly ill-posed and much information is lacked on the imaging process. To recover physically correct information, the analysis of the light rays is important.

Separating Layers from Superposed Images

Our work aims at recovering the appearance of inner slices from superposed observation. There are threads of works for layer decomposition from super-imposed images [74, 75, 76]. Since these methods are designed for semi-transparent layers that do not exhibit multiple scattering, the image formation model is represented by simple alpha blending [77]. Our method, on the other hand, is developed for dealing with translucent scenes where significant multiple scattering is observed, where alpha blending cannot describe the image formation.

In computer graphics area, various image formation models for representing subsurface scattering have been proposed. Donner *et al*. [78] propose a multi-pole model for rendering translucent layered objects such as human skin by extending dipole [79] and Kubelka-Munk theory [80]. d'Eon *et al*. [81] propose an accurate image formation model for translucent objects by decoupling single and multiple scatterings and quantizing the diffusion Green's function [82]. Based on [81], Jakob *et al*. [83] propose a framework of rendering arbitrarily layered materials. These models enable high-quality rendering of translucent objects with multiple scattering. Since they are designed for the forward rendering process, it is not straightforward to directly adopt them in our backward inverse problem, *i.e.*, recovering appearance of inner layers of translucent objects, due to the computational complexity. In this work, we use a simplified version of Kubelka-Munk theory [80] for the image formation model, which makes the inverse estimation problem tractable while retaining the expression ability.

There have been independent developments of technologies for imaging internal structures of target objects for special purposes. In art analysis, several techniques have been developed for imaging hidden layers of paintings. Infrared reflectography [84] and X-ray transmission radiography [85] have been used for visualizing internal layers of paintings, although the surface texture cannot be separated. X-ray fluorescence technique [85] uses spectroscopic information measured over a couple of days and estimates the metallic atom distribution for determining

colored appearance of inner layers. Tera-hertz imaging [86, 87] is another technique that can see inner layers of paints and books. In the medical imaging and its related areas, optical coherence tomography [88, 89] techniques are widely used, especially for visualizing retina. These techniques enable to observe inside translucent objects based on interference of coherent light at the cost of expensive wavelength-order optics and mechanics. In contrast, our method uses a commodity camera and projector for recovering slices inside translucent objects, which allows low-cost implementation. In microscopy, there are methods that use pattern projection for visualizing inside substances, or reconstructing 3-D shape of a small sample, such as protein structure [90, 91, 92]. They sharpen the microscopic image by precisely taking into account both micro-scale optical blur and scattering blur. Our aim is to develop a technique that is applicable to a standard scale, where scattering blur becomes more significant.

Time domain coding for analyzing light transport is another approach to recover images that are not directly measurable. Heide *et al*. [45] sweep the modulation frequency and phase of their customized Time-of-Flight (ToF) camera to recover the light propagation inside scattering medium. Kadambi *et al*. [39] build a coded-illumination ToF camera with a deconvolution technique and use it for recovering a sharp image by observing through a diffuser. O'Toole *et al*. [40] combine spatial probing and ToF imaging to separate direct and indirect light-in-flight images. Tadano *et al*. [62] propose an imaging system that is capable to select a target depth using a coded ToF camera. While these methods effectively recover light transport, they require carefully tailored ToF cameras. Contrary, our method uses a simple projector-camera system and spatial pattern coding to analyze light transport inside translucent objects.

Recovering Transparent and Translucent Shape

Methods for measuring transparent or translucent media, such as smoke scanning, are also related to our work. Morris *et al*. [93] recover the shape of clear transparent objects that refract light by recording light rays from different viewpoints. Hawkins *et al*. [94] acquire the density distribution of participating media, such as smoke, by laser scanning. Ihrke and Magnor [95] reconstruct the volume of dynamic semi-transparent media by a visual-hull based method. Inoshita *et al*. [96, 2] reconstruct the 3-D shape of both thin and thick translucent objects by tracing single scattering rays and assuming surface normal convolution model, respectively. While related, we recover slices inside translucent objects instead of reconstructing 3-D shape of an object's surface in Chapter 4. We also improve the shape of heterogeneous translucent object by examplar-based approach using Time-of-Flight distortion described in Chapter 6.

Similar to our work, Shim and Lee [53] reconstruct the shape of thin *translucent* objects using a single ToF camera. They model the distortion of the depth by two layer (2-sparse) multi-path model, and determined object and background layer's depth from multiple observations of different illumination phase delay. Unlike their method, our method is developed for recovering shape of *transparent* objects in Chapter 3, which have two refraction points and considerable thickness, using the variation of speed of light inside the object and refractive paths.

Early works of recovering the shape of transparent objects include [97], which recovers a single refraction water surface by observing the image placed under water. The setting of single refraction scenes has been further studied by several researchers. Morris and Kutulakos [98] reconstruct the shape of a dynamic wavy surface by observing a reference pattern placed under water from stereo cameras. Alterman *et al*. [99] estimate the position of the target object in the air from a camera placed in water using a stereo image sequence. Tian and Narasimhan [100] simultaneously estimate the shape of water surface and planar underwater scene from an image sequence by water surface tracking based on the distortion model using the wave equation. Tian and Narasimhan [101] also remove distortion of wavy surface in a data-driven approach and reconstruct wavy surface by spatially integrating the water distortion. Wetzstein *et al*. [66] reconstruct thin transparent objects by assuming thin transparent objects as a single refraction surface using light-field probe, which converts the position and angle of the light source into color codes. In contrast to these approaches that assume a single refraction path, our method focuses on scenes with two refraction paths to estimate whole shape.

There are also methods that analyze solid transparent shapes that exhibit two or more refractions. Kutulakos and Steger [102] show a general theory of tractability of shape recovery based on refractive paths characterized by the number of viewpoints, reference points, and refraction points. They also show reconstruction of a transparent shape from three viewpoints, two reference points, and two refraction points. Their problem setting is similar to ours but our method reconstructs from two observations from a single viewpoint using a ToF camera, with moving the background reference surface. There are also some single viewpoint approaches for transparent shape reconstruction. Morris and Kutulakos [93] reconstruct inhomogeneous transparent objects by illuminating an object from various positions and analyzing the specular reflections. Similarly, Yeung *et al*. [103] also propose a reconstruction method based on specular reflection analysis. While effective, these methods require a number of light positions for accurate recovery.

Some other methods take an approach of using specialized devices for recovering transparent shape. Han *et al*. [104] reconstruct a single side of transparent surface object by the

	# view	# ref.	specialized device
[102]	3	2	–
[93, 103]	1	many	–
[66]	1	1 (+angle)	light-field probe
[104]	1	4	liquids
Ours	1	2	(retroreflective screen)

Table 2.1: Relation to the prior work. Our method recovers transparent shape from single viewpoint and two reference points without specialized devices.

crossing of two reference rays, one of which is measured in the air and the other is measured in liquid. There are also other unique methods for recovering transparent surfaces. Ihrke *et al.*[105] reconstruct the shape of flowing water by dyeing water with a fluorescent chemical and observing from multiple video cameras. Ma *et al.*[106] acquire the refractive index field based on the transport of intensity equations, which is a theory of phase imaging with coherent illumination, using collimated illumination, and reconstruct 3-d refractive volume based on tomography. Miyazaki and Ikeuchi [107] propose an inverse polarization ray-tracing for estimating the front surface of a transparent object using polarized reflections. Eren *et al.*[108] use thermal imaging for determining transparent shape by illuminating the target by laser beam.

Table 2.1 summarizes the settings of transparent shape estimation methods, mainly single viewpoint approaches. Our method determines the shape of transparent objects from a single viewpoint with two reference points. We consider that a ToF camera is now a commodity device because it is available at a similar cost with ordinary RGB cameras.

Similar to our setting, there are works that use a ToF camera with scenes including transparent objects. Heide *et al.* [36] recover light propagation sequences of a scene by sweeping the modulation frequency of their custom ToF camera. O'Toole *et al.* [40] separate light-in-flight images into direct reflection, specular inter-reflection, and global components such as caustics based on spatial probing. Kadambi *et al.* [39] reconstruct time sequential images of a scene including transparent objects by altering ToF measurement using coded light pulses. Gkioulekas *et al.* [31] visualize propagation of light by an optical coherence tomography scheme using a Michelson interferometer. Because their method has pico-second time resolution, spectral dispersion in a glass slab can be visualized. While these works visualize light transport of the scene, our goal is reconstruct the shape of transparent objects.

Material Estimation

Non-invasive and non-contact material classification is an important research topic in computer vision and yet remains a challenging task. There are several prior works for material estima-

tion. The methods based on the visual appearance, *e.g.*, color, shape, and/or textures of the material [109, 110, 111, 112, 113, 114, 115, 116], typically only require a single RGB image; thus, the setups are easy to realize. The main problem is that this approach suffers from similar appearances of different materials, *e.g.*, texture-less boards, resulting in a lower accuracy due to the lack of information.

The class of approaches based on the optical properties, such as BRDF [117, 118], shading [119], and spectrum [120], has a capability of distinguishing visually similar objects in higher accuracy because the optical properties convey richer information about the material. However, constructing such measurement systems and building database of samples generally require carefully controlled settings. This class includes approaches based on other physical properties, *e.g.*, elasticity [121], and water permutation and heating/cooling process [122]. Our method falls into this class because we use a temporal response of the incident light, which implicitly measures the optical and physical properties of target objects. Unlike these approaches, our method uses an off-the-shelf ToF camera and needs only single observation at least, hence the cost of constructing the system is as low as the appearance-based methods.

In the context of material classification using a ToF camera, Su *et al.*'s method [123] is closely related. They propose a method that classifies a material from raw ToF measurements by sweeping over several modulation frequencies and phases. While the approach is shown effective, it requires special customization of a ToF camera for obtaining the measurements. In contrast, our method only uses an off-the-shelf ToF camera. We show that the material classification can be achieved by such a simple setup by exploiting the depth-dependency of the measurements.

Chapter 3

Light Path Analysis of High-frequency Illumination

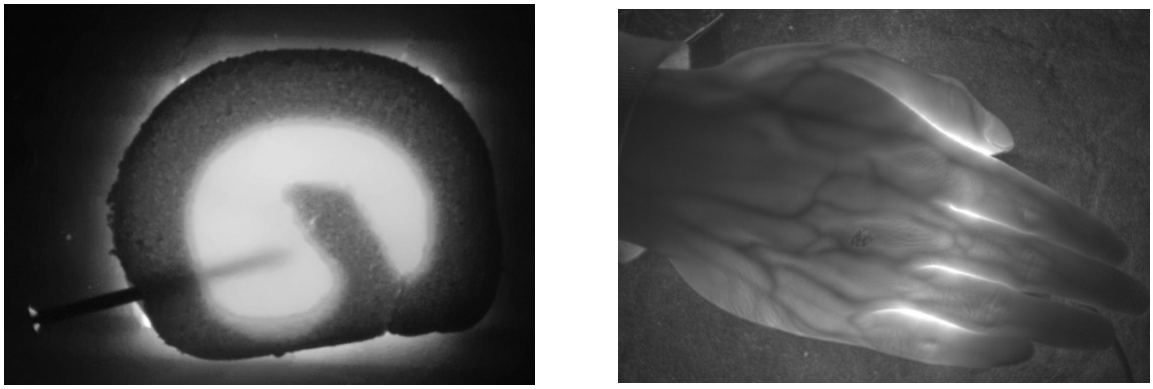
In this chapter, we show the spatially modulated illumination techniques. Specifically, we use a projector to illuminate the scene by spatially amplitude-modulated light, called 'high-frequency illumination'. We show that the light path design is important for high-frequency illumination. Applying this theory for transmissive observation, we recover clear images of the scene through scattering media.

3.1 Preliminary

When recording transmissive observations of a translucent object, captured images tend to be unclear due to strong scattering. The scattering is caused by collisions of light rays with tiny particles in the medium that diffuse light rays, and is inevitable in various translucent objects. Therefore, for recording a clear transmissive images, an effective descattering technique is needed.

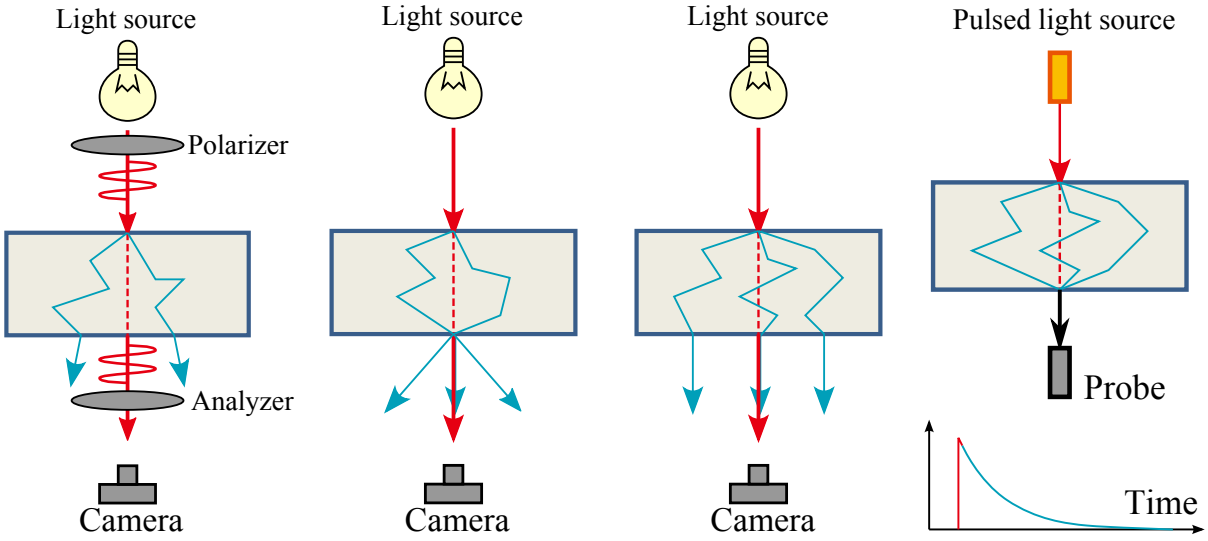
Figure 3.1 shows examples of transmissive images. A nail in a piece of food is shown in (a). There is an urgent demand in the food industry to inspect products for detecting and eliminating foreign objects, where transmissive imaging can help. Another example as shown in (b) shows vein patterns inside a hand. Transmissive imaging is also useful for living body visualization, measurement of blood oxygen levels [63], and security purposes. In such application domains, capability of recording clear transmissive images plays an important role.

In this chapter, we propose a *Parallel High-frequency Illumination* method which sharpens transmissive images by separating the transmissive and scattered lights. We show in this chapter that transmissive and scattered lights can be separated based on angular and positional clues. We then show that the original high-frequency illumination method proposed by Nayar *et*



(a) Foreign object in food (b) Vein pattern

Figure 3.1: Examples of transmissive image. Although obstacles inside the object can be visualized, the image is blurred and unclear because of multiple scattering.



(a) Polarization (b) Angle (c) Position (d) Time

Figure 3.2: Differences between transmissive and scattered lights. Direct tramsimssive light and indirect scattered lights exhibit a different optical properties with respect to polarization, emission angle, spatial spread, and time domain impulse response.

al. [12] and several extended methods [13, 14, 15] can be uniformly explained as a separation of the overlapped and non-overlapped rays. We finally evaluate the effectiveness of the proposed parallel projection approach to separating transmissive and scattered lights.

3.2 Parallel High-frequency Illumination

Before introducing our method, we start from the transmissive observation in the point of the difference between direct transmissive light and scattered light.

Transmissive Images

Transmissive image such as those in Fig. 3.1 can be obtained by placing a camera and a light source on opposite sides of the translucent object. The lights observed by the camera contain two components. One is the *transmissive light*, which passes straight through the scattering object. The other is the *scattered light*, which has various path changes caused by collisions with tiny particles. While the former component shows the internal structures in a translucent object directly, the latter component disturbs the clarity of the observation. To obtain a clear transmissive image, we need to eliminate the scattering lights and extract only the transmissive lights.

Before separating the transmissive and scattered lights, we must clarify the differences between the two components and determine the clues that can be used for our purposes. We summarize several clues for separation of the two components and introduce some conventional methods in the following.

Polarization clue

When the incident light is polarized, the transmissive and scattered lights have different polarization characteristics, as shown in Fig. 3.2(a). While transmissive lights maintain their polarization because they are not affected by the object, the polarization can be gradually lost because of the collisions with tiny particles. Therefore, scattered lights which have bounced several times become only partially polarized, and the polarization is perfectly lost after multiple repeated bounces. Based on this difference, the scattered light can be reduced by the use of two co-linear polarizers [68].

Angle clue

Lights are emitted from an observed point to various directions as shown in Fig. 3.2(b). This emission angle is different for the transmissive and scattered lights. Because the path of the transmissive light does not change when passing through the object, the emitted light has the same direction as the incident light. In contrast, the scattered light's path variously changes by collisions, and thus the emission direction also changes. Based on this difference, Kim *et al*. [69] have estimated the intensity of the transmissive light from angular observation using a light field camera. Brogili *et al*. [65] have applied the schlieren method, which obscures the transmitted ray by using a knife edge, to measure the scattering lights.

Position clue

Scattered and transmissive lights have difference about emission positions as shown in Fig. 3.2(c). While the transmissive lights are emitted on the same line as the incident light, the scattered light is emitted from various positions because they are spread by the scattering particles. If the incident light is a narrow beam, the transmissive light is observed as a pointed peak. In such a case, when the incident point of the beam shifts slightly, the position of the observed transmissive light also shifts. Otherwise, the scattered lights spread smoothly, and a slight shift in the incident beam does not affect to the observation. The positional cue is one of our key observations for descattering, and we use a pattern projection method for separating spatially spread light.

Other clues

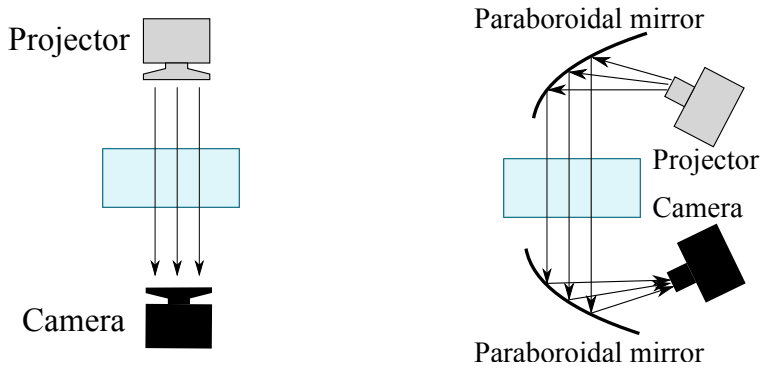
Other clues have been also used for the separation. For example, the path length of scattered light rays obviously becomes longer than that of transmissive light rays, therefore the transmissive lights arrive earlier than the scattered lights as shown in Fig. 3.2(d). By using an ultra high speed sensor such as a streak tube, the two components can be separated based on this time clue [58]. A wavelength clue can also be used. Because the scattering and absorption are wavelength-dependent effects, multi-band illumination can be used for deblurring [124].

Parallel projection system

Our aim is to sharpen the transmissive images effectively based on the differences between the transmissive and scattered lights. As explained above, there are several clues that can help to detect the separation of these lights. Unfortunately, there is no single perfect way. For example, the angle clue cannot be used to distinguish transmissive lights from straight-line scattered lights. Similarly, the position clue cannot be used to distinguish transmissive lights from scattered lights that are emitted from the same position.

To overcome this problem, we use a combination of both the angle and the position clues. To combine the different types of clues, we have designed a special optical system that can illuminate and observe high-frequency patterns *in parallel*, as shown in Fig. 3.3 (a). This system can also easily combine the polarization and wavelength clues by using a polarizer and a near infrared light source, and thus we can use many differences as possible.

Two optical designs are available to realize the parallel projection. The first way is to use telecentric lenses as shown in Fig. 3.3(a). Although telecentric lenses are expensive and the



(a) Telecentric lenses (b) Paraboloidal mirrors

Figure 3.3: Optical design for parallel projection. Parallel light rays can be generated using telecentric lenses and putting the camera where the focal points of both the camera and the paraboroidal mirror coincide.

viewing area becomes narrow, a variety of lenses are commercially available and this design is easy to configure. The other design is to use paraboloidal mirrors as shown in Fig. 3.3(b). Although paraboloidal mirrors are low cost and the viewing area can easily be extended, it is necessary to precisely align the paraboloid's focal point to the optical centers of both the camera and the projector. In practice, we have to select the design according to the size of the scene, the cost, and the difficulty of the setup.

Decomposition of transmissive and scattered lights

Let us assume that $L_o(q)$ is the light intensity observed by a camera at pixel q. L_o can be represented as

$$L_o(q) = L_t(q) + L_s(q), \tag{3.1}$$

where $L_t(q)$ and $L_s(q)$ are the intensities of the transmissive and scattered lights at pixel q of the camera respectively.

Our method separates the transmissive and scattered lights by projecting effectively small checkered pattern, a similar way to the original method [12], but in parallel. When the phase of the projected pattern changes slightly, L_t also changes, but L_s rarely changes. Therefore, we can separate the light components based on this difference. We assume that the numbers of white and black pixels of the projected pattern are same. When the projected pattern is shifted, the maximum intensity $L_{\max}(q)$ and the minimum intensity $L_{\min}(q)$ of all observations

are represented as

$$L_{\max}(q) = L_t(q) + \frac{1}{2}L_s(q), \tag{3.2}$$

$$L_{\min}(q) = \frac{1}{2}L_s(q). \tag{3.3}$$

From these relations, the transmissive and scattered lights can be computed as

$$L_t(q) = L_{\max}(q) - L_{\min}(q), \tag{3.4}$$

$$L_s(q) = 2L_{\min}(q). \tag{3.5}$$

We call this separation method the *Parallel High-frequency Illumination* method.

3.3 Light Path Design for High-frequency Illumination

Nayar *et al*. [12] have proposed the original high-frequency illumination method that separates the direct and global components. However, the photometric factors that directly correspond to the direct and global components depend on the settings of illumination and observation. Every separation method has a different scheme for pattern projection. The separable components are also different. However, we found that these methods share a commonality. In fact, these methods can be uniformly defined as a method for separation of overlapped and non-overlapped light rays. If we need to separate certain components, then we only have to design a non-overlapping projection system for these components.

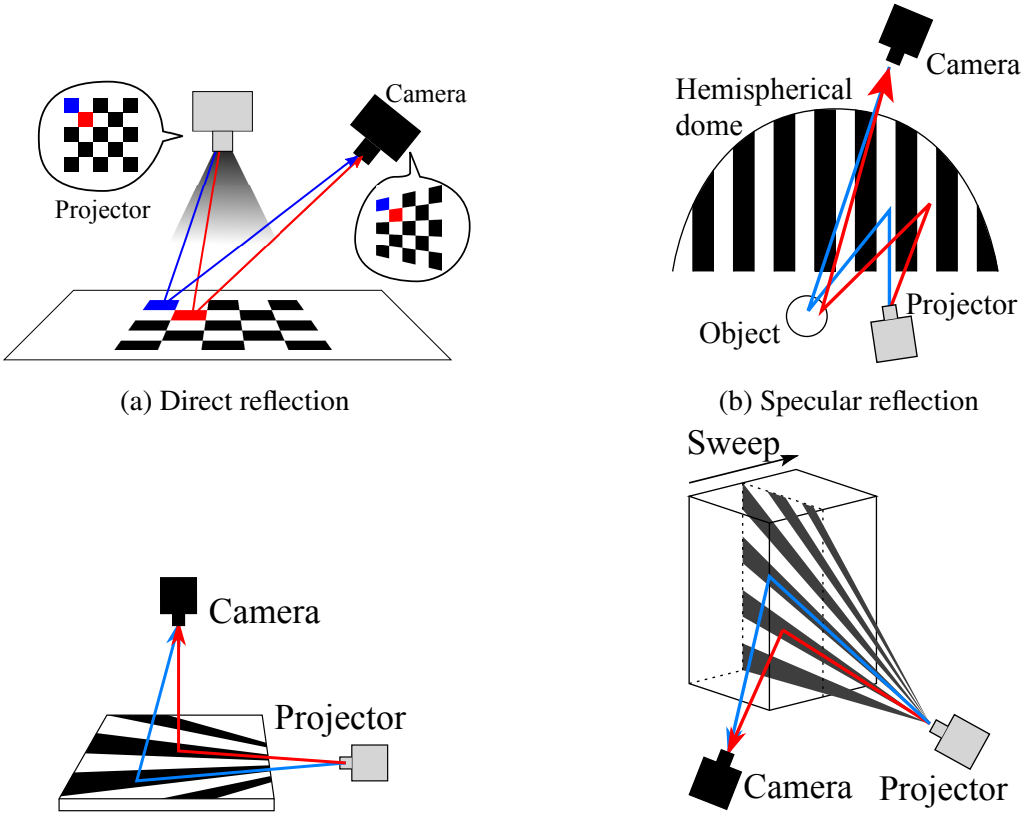
First, we show how the direct and global components change when the illumination and observation settings change in normal perspective projection.

Separation of direct and indirect reflections

The direct components in the original method [12] are the diffuse and specular reflection, while the global components are inter-reflection, subsurface scattering, and volumetric scattering. The red line in Fig. 3.4(a) is an example of light path from the projector to the camera. This red line does not overlap with the other paths indicated by the blue line. Because these light paths do not overlap with each other, the high frequency components remain.

Separation of diffuse and specular reflections

Lamond *et al*. [13] show a separation method for diffuse and specular reflections. They projected a stripe pattern on to a hemispherical screen surrounding the object and observed its reflection, as shown in Fig. 3.4(b). In this case, the direct component corresponds to the specular reflection and the rays do not overlap with each other.



(a) Direct reflection

(b) Specular reflection

(c) Planar single scattering

(d) Volumetric single scattering

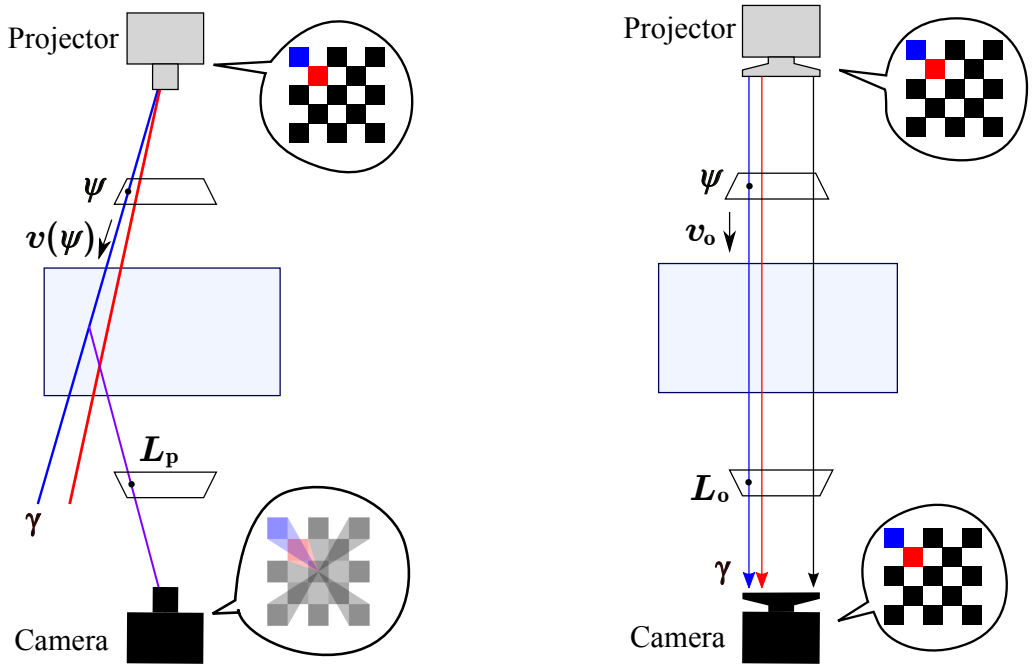
Figure 3.4: Light extraction by high-frequency illumination. Careful optical design realizes different light extraction, where extracted lights do not overlap each other.

Separation of single and multiple scatterings in a plane

Mukaigawa *et al*. [14] have separated single and multiple scattering components in a thin planar translucent object. They projected a stripe pattern from the side of the object and observed vertically, as shown in Fig. 3.4(c). In this case, the direct component corresponds to the single scattering. The important point here is that the target is limited to a plane, and thus each ray is non-overlapped.

Separation of single and multiple scatterings in a volume

Mukaigawa *et al*. [15] have extended their previous method for volumetric object. They swept the stripe pattern and observed each depth independently, as shown in Fig. 3.4(d). In this case, the direct component corresponds to the single scattering. If a two-dimensional pattern is projected on a volumetric scene, the single scattering light rays are overlapped in the captured image. This method avoids the overlapping problem by observing each depth independently.



(a) Perspective projection

(b) Parallel projection

Figure 3.5: High frequency illumination for transmissive observation. Because lights are overlapped on the camera plane, perspective projection is not suited for transmissive observation. On the other hand, the projector's pixel and camera's pixel have one to one correspondence hence the direct transmission can be separated.

3.3.1 Pixel Correspondence on Parallel Projection System

Let us assume a simple setting in which a projector and a camera are placed face to face, as shown in Fig. 3.5(a). If a normal projector and a normal camera are used for high frequency illumination, the transmissive rays cannot be extracted as a direct component. This is because the transmissive rays are overlapped. Let us consider a light ray $\boldsymbol{\gamma}$, which corresponds to a pixel $\boldsymbol{\psi}$ in the projection pattern. This ray can be expressed as

$$\boldsymbol{\gamma} = \boldsymbol{\psi} + j\boldsymbol{v}(\boldsymbol{\psi}), \tag{3.6}$$

where $\boldsymbol{v}(\boldsymbol{\psi})$ denotes the direction of the ray, which depends on the projector's pixel $\boldsymbol{\psi}$, and j is an arbitrary parameter. This ray is captured by a camera, and the captured image $\boldsymbol{L}_p$ is expressed as

$$\boldsymbol{L}_p = \mathbf{P}(\boldsymbol{\psi} + j\boldsymbol{v}(\boldsymbol{\psi})), \tag{3.7}$$

where $\mathbf{P}$ denotes the perspective projection matrix of the camera. This shows that a pixel in the projected pattern is observed as a line in the captured image. Therefore, there is no one-to-one

correspondence between the projected pattern and the captured image. For example, red and blue lines in the projected pattern produce two rays which do not overlap in the 3-D volume, as shown in Fig. 3.5(a). However, the two rays overlap with each other in the image captured by the camera.

On the other hand, parallel projection produces different relationships. Let us assume that the projector and the camera are arranged to produce a coaxial orthographic projection, as shown in Fig. 3.5(b). In this case, a ray $\boldsymbol{\gamma}$ can be expressed as

$$\boldsymbol{\gamma} = \boldsymbol{\psi} + j\boldsymbol{v}_o, \tag{3.8}$$

where $\boldsymbol{v}_o$ denotes the direction of the ray, which is the same as the projection direction. It should be noted that $\boldsymbol{v}_o$ does not depend on the pixel $\boldsymbol{\psi}$. The coordinates in the captured image are expressed as

$$\begin{aligned} \boldsymbol{L}_o &= \mathbf{O}(\boldsymbol{\psi} + j\boldsymbol{v}_o) \\ &= \mathbf{O}\boldsymbol{\psi} + j\mathbf{O}\boldsymbol{v}_o, \end{aligned}$$

where $\mathbf{O}$ denotes the orthographic projection matrix of the camera. Here, $\mathbf{O}\boldsymbol{v}_o = 0$ because $\boldsymbol{v}_o$ is degenerated by the projection. Therefore,

$$\boldsymbol{L}_o = \mathbf{O}\boldsymbol{\psi}. \tag{3.9}$$

This shows that the ray is projected as a point in the captured image. For example, the red and blue pixels produce two parallel rays, and they are observed as being different points without overlapping. Therefore, the transmissive and scattered lights can be separated as non overlapped and overlapped rays, respectively. This is the core essence of our *Parallel High-frequency Illumination* method, which is based on both the angle and position clues.

3.4 Experiments

Measuring system We have constructed a parallel projection system using telecentric lenses, as shown in Fig. 3.6. We have used the digital micro-mirror device (DMD) projector development kit (Texas Instruments LightCommander) with a telecentric lens (Edmund Optics). This projector has an infrared (850 nm) light source along with visible RGB light sources. The CCD camera (Point Grey Grasshopper2) which has sensitivity to the near infrared light, also has a mounted telecentric lens. Linear polarizers are placed in front of both lenses to reduce scattered lights. Using this equipment, we have realized parallel high-frequency illumination in the near

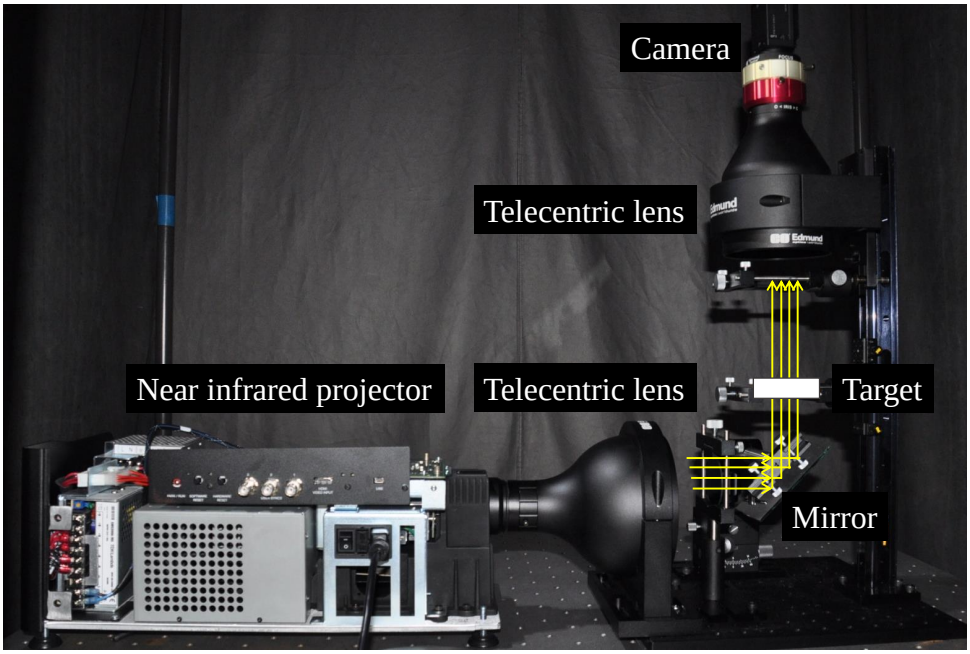


Figure 3.6: Parallel projection system. We use an infrared projector and camera, and parallel projection is realized by confronting telecentric lenses.

infrared region. For the high-frequency illumination, a 9×9 pixel checker pattern is projected. We shift the pattern for one sixth phase and capture 18 images. This size provides a threshold to determine the allowed spread of the transmissive lights. The appropriate size depends on the object size, the resolutions of both the camera and the projector, and the optical density.

Alignment Because parallel projection has high sensitivity about light ray's angle, both the camera and projector must be aligned their optical axes. To confirm the alignment, we capture a thin diffuser several times at different height and confirm they are exactly the same image. We check the difference of images while projecting high frequency pattern to be zero.

Experiment with diluted milk First, we have confirmed the descattering effect by the proposed method. The target object is a metallic wire, as shown in Fig. 3.7 (a). We regard a clear transmissive image which is captured in pure water as the ground truth, as shown in Fig. 3.7(b). The blurry scene is produced by pouring a small amount of milk into the water. The density of the diluted milk is set to 2.2%, where the target object is hardly seen in a normal observation.

We use two wavelengths of visible (525 nm) and near infrared (850 nm) light for comparison purposes. Figure 3.7 (c, d, e) show the experimental results. The image as shown in Fig. 3.7(c) shows a normal transmissive image in visible light. Because the light is strongly scattered, it is hard to recognize the object. The image as shown in Fig. 3.7(d) is a normal

(a) Target object (b) Target object in pure water (c) Visible light (d) Infrared (e) Our method

Figure 3.7: Experimental results. (a) the target object. A metallic spring is sunk in dirty water. (b) The target object is sunk in pure water. Clear image can be observed. (c) A normal observation is unclear because of scattering. (d) Infrared light improve the image, however it is still low contrast. (e) Our result. Clear observation is recovered.

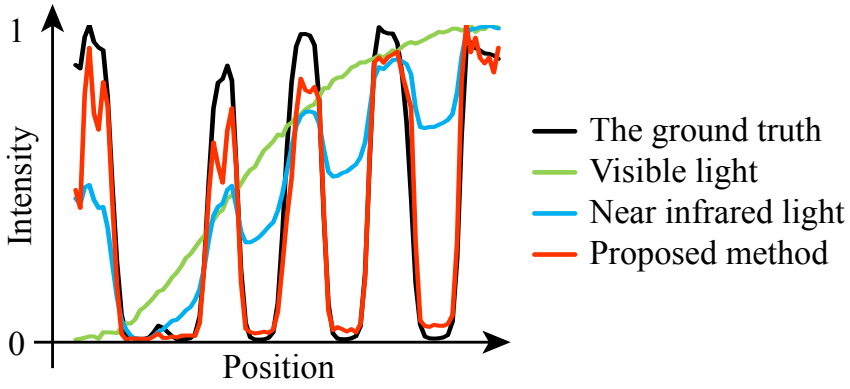


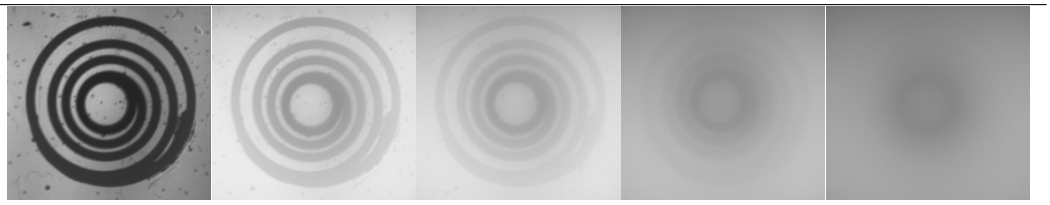
Figure 3.8: Comparison of intensity profiles. Intensities along with the red line as shown in Fig. 3.7 (b) is plotted.

Table 3.1: Normalized cross correlation. Value ranges between -1 to 1, where perfectly correlated images are 1.

Normal illumination in visible light	Normal illumination in near infrared	Proposed method
-0.10	0.68	0.95

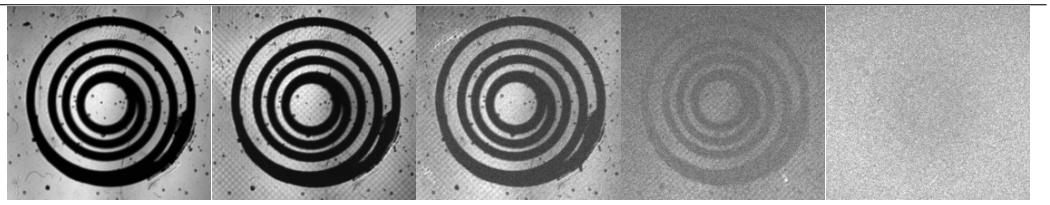
transmissive image using the near infrared light. The use of the near infrared light contributes to sharpening of the image, but it is insufficient and the contrast is low. The image as shown in Fig. 3.7(e) shows the result of our proposed method. This confirms the image sharpening effect while it contains some artifacts due to pattern projection. Figure 3.8 shows a comparison of the intensity profile of each image along with the red line as drawn in Fig. 3.7 (b). The image from the proposed method has sharp edges and its profile is similar to the ground truth profile. Table 3.1 shows the result of normalized cross correlation with the ground truth image. The proposed method produces the highest value, which confirms the nobility of it.

Next, we confirm the descattering effect for various milk densities. we use 1.9%, 2.2%,



density	1.9%	2.2%	2.5%	2.8%	3.1%
correlation	0.98	0.68	0.30	0.06	0.01

(a) Normal illumination of near infrared



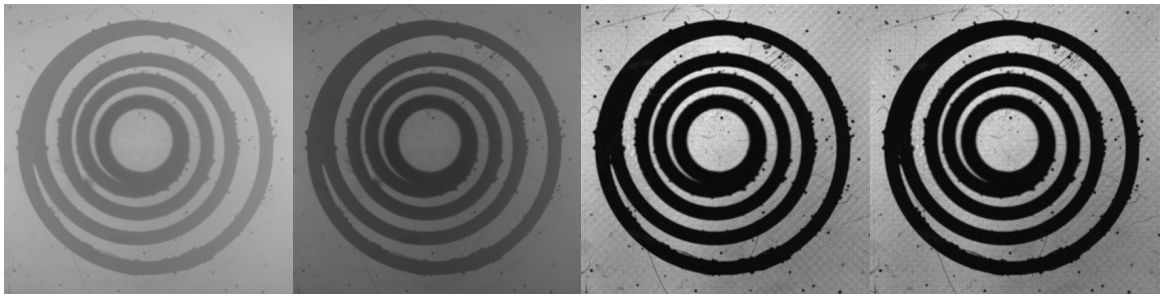
density	1.9%	2.2%	2.5%	2.8%	3.1%
correlation	0.95	0.95	0.84	0.21	0.03

(b) Proposed method

Figure 3.9: Results for various densities. Our method recovers clearer images under 2.5% density, and is limited for higher density.

2.5%, 2.8%, and 3.1% milk density values for our experiment. Figure 3.9 compares the results of normal illumination with near infrared light and the proposed method. The normalized cross correlation with the ground truth is computed for each image. As the density of the milk increases, clarity of the transmissive images decreases because scattering blur becomes stronger. All images between 1.9% and 2.8% are clearly sharpened by the proposed method and the correlation value increases. At 1.9% density, the normal image is still clear but the result is actually worse, because it contains noise caused by the pattern projection process. On the other hand, for 3.1% density, it is difficult to confirm a good result because the level of the transmissive light is the same as the level of the camera noise.

Performance of the combination with polarization analysis While polarizers used in our system are fixed to co-linear for our experiments, our method can be combined with traditional descattering methods using polarization cue. That is, the difference of two observations captured in co-linear and cross settings is used to reduce scattering because scattering lose the polarization property. Capturing both co-linear and cross settings for each pattern projection, the input of high-frequency illumination is improved thus the result may be more improved. Figure 3.10 is a comparison of our baseline method and the combined method. The effect of the polarization analysis is much smaller than our proposed method because the correlation



0.916	0.937	0.988	0.988
(a)	(b)	(c)	(d)

Figure 3.10: Comparison with polarization analysis. Values under the images show cross-correlations with the ground truth. (a) Normal illumination (b) Traditional polarization analysis (c) Our proposed method (fixed polarizers) (d) Combined method

value is the same to the third decimal space. Although the polarization analysis can be easily combined, the quality is not improved and measuring time becomes two times longer.

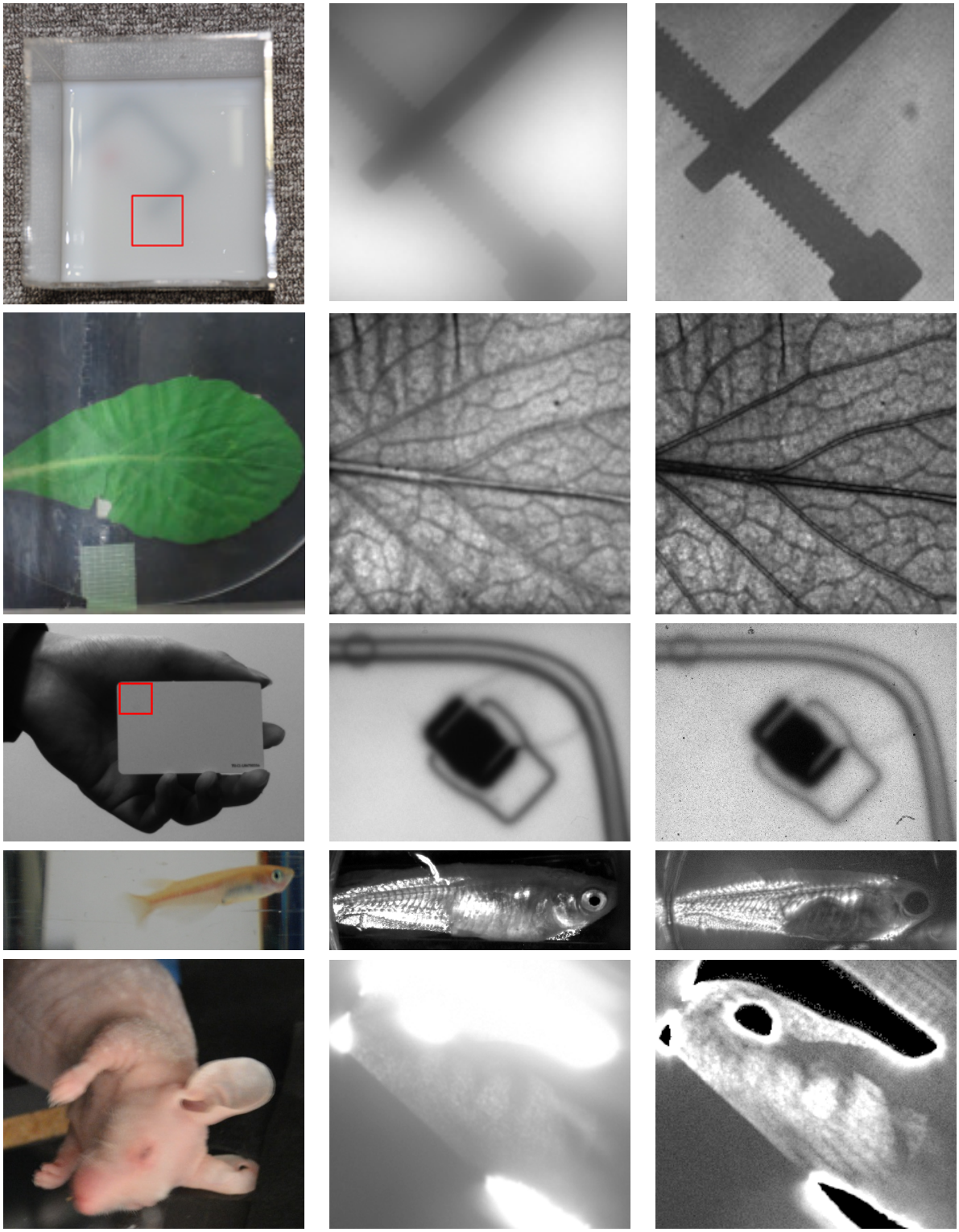
Experiments with several scenes Finally, we apply our method to several scenes, as shown in Fig. 3.11. For the scene where a screw is sunk in the dirty water, the clear image can be recovered so that the detail of the pitch can be observed. The vein pattern is observed clearly for the image of the vein of a leaf, and the circuit of the IC card is clarified. Our method can be also applied for living body observations. Backbone and the inner organ of a small fish is recovered by our method, while they are invisible in a normal infrared image of a fish. The last row shows that the inner structure of the ear of a mouse is also visualized. Applicability to the living body is confirmed by these experiments.

3.5 Discussions

In this chapter, we have proposed a descattering method of transmissive observation, where the transmissive and scattered lights are separated by Parallel High-frequency Illumination. We have explained the principles of the high-frequency illumination method and the importance of avoiding overlapping rays. We have proposed an optical design for transmissive observation so that the correspondence of the projector and the camera is preserved. We have developed a projector-camera system using telecentric lenses for our proposed method, and confirmed the effectiveness of the sharpening process by several real-world experiments, numerical evaluations, and living-body applications. While we have confirmed that the proposed method is effective for sharpening of transmissive images via several experiments, some limitations become clear.

Firstly, transmissive observation is sometimes suffered from a large amount of noise. Especially an optically thick translucent object, the ratio of transmissive light to the scattered light becomes quite small. It means that the transmissive light buries in the observation noise, and is difficult to recover. Using a denoising techniques and using a low-noise camera such as a cooled CCD camera can mitigate this problem, it is still difficult to solve essentially. Related to this problem, Our method cannot be applied to the scene where there are no transmissive lights. It is therefore impossible to visualize the internal organs of the human body using near infrared light.

Secondly, the results are degraded if the object surface is not planar or parallel to the image plane. When the surface is not a parallel plane, the light ray refracts on the surface and breaks their parallel path, therefore high-frequency illumination cannot be applied. We must have to devise a method to prevent refraction, *e.g.*, index matching by submerging the target object in a liquid with the same refractive index as the object.



Target objects | infrared image | Proposed method

Figure 3.11: Results for several different scenes. A screw inside diluted milk, a leaf, an IC card, a small fish, and the ear of a mouse are measured. While normal observations using infrared light are blurred and low-contrast, our method produces clear direct images.

Chapter 4

Analysis of Spatially Spread Lights and Multi-frequency Illumination

In this chapter, we propose a method for recovering inner slices of translucent objects such as oil paints. The key observation for this problem is that the scattering lights inside translucent object spread greater along with the depth. We use this cue for separating slices of inner layers by projecting varying frequencies of high-frequency illumination. We analyze the relationship between the different frequencies of illumination and depth dependent PSFs, and recover clear images of the inner layer of translucent objects.

4.1 Preliminary

Seeing inside a translucent object is a difficult task because its appearance is superposition of light rays emitted from every inner layers that are blurred due to subsurface scattering. Because observing the internal appearance of objects is of broad interest in medical, art, and industrial inspections, various imaging techniques have been developed for achieving this goal in the past. In particular, since the translucency effect becomes significant for many materials in near infrared (NIR) wavelengths, infrared photography is used as one of common techniques for this purpose. For example, it has been used for observing inner layers of oil paintings that tell us a lot about the drawing technique, history, and/or authenticity of old-age painters.

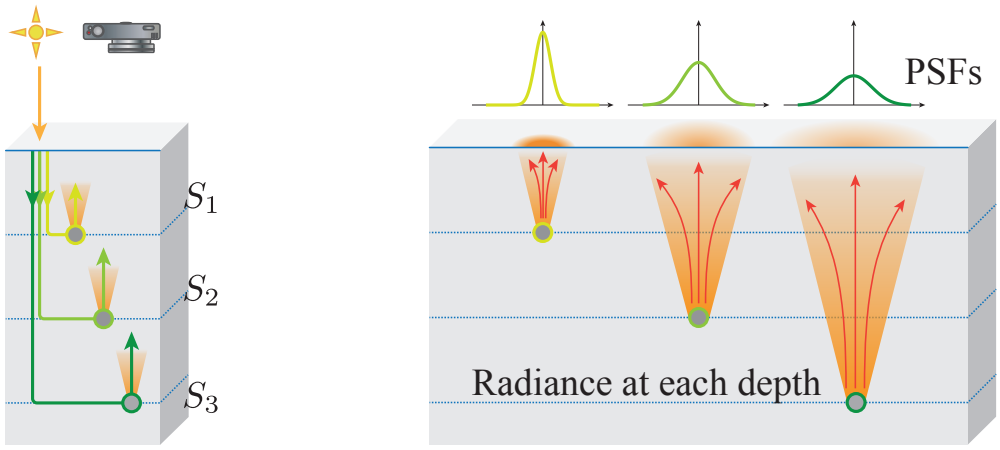
One of the major difficulties in observing inner layers of translucent objects is to separate individual inner appearances with properly dealing with scattering. To overcome this difficulty, we develop a *multi-frequency illumination* method, which can recover sharp appearance of inner slices at a desired depth with explicitly removing scattering blur. Compared with conventional techniques that aim at a similar goal, our method is faster and safer than the X-ray fluorescence technique [85], and sharper results can be obtained differently than infrared reflectography [84].

Our method exploits the fact that the spread of light due to scattering has dependency on the depth of inner layer from which light rays are emitted. By modeling the light spreads as depth-dependent point spread functions (PSFs) and exploiting their depth-dependent low-pass characteristics, we develop a method for recovering inner layer appearances from a set of images taken under a variant of high-frequency illumination [12]. Specifically, our method uses a spatial pattern projection with varying its pattern pitch – we call it *multi-frequency illumination*. Our multi-frequency illumination method allows us to separate direct (high-frequency) and global (low-frequency) components in a similar manner to [12], yet at various frequency levels that define high- and low-frequencies. The observed direction components are then used for recovering the appearance of inner slices, which are related to the direct components via depth-dependent PSFs.

Furthermore, to deal with inhomogeneous translucent objects that exhibit spatial PSF variations due to its inhomogeneity, we develop an extended method for determining the pixel-wise PSFs with an assumption of spatial smoothness of the target depth. In addition, to reduce the visible shadowing effect that appears in the inner layers, which is due to occlusion by their upper layers, we develop a shadow detection method so that the shadowed regions can be later post-processed for better visual quality.

The key contributions of this chapter are threefold. First, we describe the relationship between depth inside a translucent object and its PSF by a physically motivated scattering model. Second, based on the relationship, we develop a method for recovering the appearance of inner slices using varying pitch pattern projection. Third, we develop a method for recovering clear inner slices of inhomogeneous objects based on the analysis of spatially non-uniform transmittance. We apply our method to the real-world scenes using a coaxial projector-camera system, and show the effectiveness of the proposed method using oil painting, mural, and paper scenes.

The rest of this chapter is organized as follows. Section 4.2 describes the image formation model for translucent objects and the depth-dependent PSF model that represents blur effects due to subsurface scattering. Based on the model, we describe the proposed method for recovering inner slices of translucent objects in Sec. 4.3. Section 4.4 shows results for both simulation and real-world data. Finally, we conclude the chapter in 4.5.



(a) (b)

Figure 4.1: Illustration of the image formation model for translucent objects. (a) Recorded intensity is the summation of all layer's appearance. (b) Spatial spread of light varies with depth.

4.2 Appearance of Translucent Objects

When an image of a translucent object is recorded, the observed intensity $L_o(q)$ at camera pixel $q \in \mathbb{Z}^2$ is a summation of all possible paths $\rho \in \mathcal{P}_z$ from depth z:

$$L_o(q) = \int_0^{\zeta} \int_{\mathcal{P}_d} \iota(\rho) d\rho dz, \tag{4.1}$$

where ζ is the maximum thickness of the object, and $\iota(\rho)$ is the intensity of the light path ρ. It can be modeled as a summation of the appearance of all depth slices as illustrated in Fig. 4.1(a), which can be described as

$$L_o(q) = \int_0^{\zeta} S_d(q) dz, \tag{4.2}$$

where S_z is the appearance slice at depth z.

The appearance slice S_z is generally blurry due to the scattering effect inside the medium. The spread of radiance at a scene point inside a translucent object varies depending on its depth from the object surface [125]. In general, the spatial spread of light can be expressed using PSFs. Let us consider light rays emitted (or returned) from a specific depth inside a translucent object. When the depth is shallower, the PSF becomes sharper. It gradually wider spreads as the depth z becomes deeper inside the medium as illustrated in Fig. 4.1(b). In this manner, there is a close relationship between the PSF and depth. By denoting h_z as a PSF at depth z, the appearance slice S_z can be expressed as

$$S_z(q) = (R_z * h_z)(q), \tag{4.3}$$

where R_z is the sharp appearance slice that we are interested in estimating, which we call a radiance slice, and $*$ denotes a convolution operator.

Since the appearance of the translucent object under normal illumination is a superposition of radiance of multiple slices as Eq. (4.2), the observation L_o can be re-written as

$$L_o(q) = \int_0^{\zeta} (R_z * h_z)(q) dz. \tag{4.4}$$

Specifically, we are interested in recovering a few planar parallel layers inside the medium; therefore, the image formation model can be discretized as a multiple layered model as

$$L_o(q) = \sum_d (R_d * h_d)(q), \tag{4.5}$$

where R_d and h_d are the radiance slice and depth-dependent PSF of d-th layer, respectively. Our goal is recovering radiance slices R_d convolved by unknown depth-dependent PSFs h_d from the composite observation L_o. Before introducing the solution method, we describe a model of depth-dependent PSFs h_d.

Depth-dependent PSFs

Our depth-dependent PSF model is motivated by the work of [126], which is a physically motivated scattering model designed for the PSF of scattering medium, named the *radiative transfer equation* (RTE) [127, 128]. In this work, we represent the depth-dependent PSFs h_d by the RTE model. With the RTE model, the intensity of an arbitrary light ray in a homogeneous scattering medium can be iteratively calculated [126] as

$$I(d, \theta) = \sum_{k=0}^{\infty} \left(g_k(T) + g_{k+1}(T)\right) L_k(\cos\theta), \tag{4.6}$$

where L_k is the Legendre polynomial of order k, $g_0 = 0$, and

$$g_k(T) = I_0 \exp\left(-\frac{2k+1}{k}(1 - g^{k-1})T - (k+1)\log T\right).$$

The parameter T $(= \sigma d)$ represents optical thickness, which is the product of scattering coefficient σ and distance d between the point light source and the scattering point. The forward scattering parameter g controls how light rays spread; $g = 0$ corresponds to isotropic, while positive and negative values of g indicate forward and back scattering, respectively. θ is the angle between the depth axis and light ray's direction as depicted in Fig. 4.2(a). k is the number of light bounces, and I_0 is the intensity of the point light source.

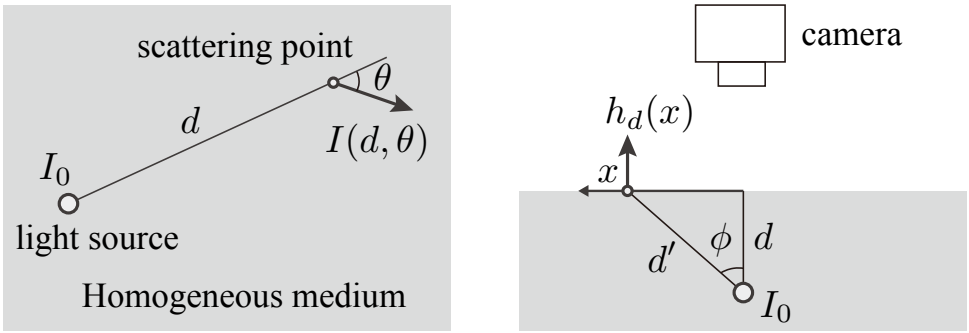


Figure 4.2: Illustration of RTE model described in [126]. (a) Solid arrow represents a scattering light ray, whose intensity depends on d and θ. d is the distance from the point light source I_0 to the point inside the scattering medium, and θ is the radial direction of the light ray. (b) Depth-dependent PSF h_d can be expressed using RTE model. We consider the x-axis lies along the tangent plane of the surface of translucent object and assume the direction of light rays emitted from the surface becomes parallel to the depth axis.

In our setting, a camera is placed outside the media, where light rays do not scatter. In this case, the depth-dependent PSF applied to a point light source inside scattering media corresponds to the intensity of the light rays emitted from the surface, if the camera is placed sufficiently far from the object. When we consider that the x-axis lies along the tangent plane of the surface of translucent object as illustrated in Fig. 4.2(b), the depth-dependent PSF $h_d(x)$ can be written as

$$h_d(x) = I\left(d', \phi\right), \tag{4.7}$$

where $d' = \sqrt{x^2 + d^2}$ and $\phi = \tan^{-1}\left(x/d\right)$. We assume that the camera is placed sufficiently far from the object compared to both d and x so that the direction of emitting light rays from the surface becomes parallel to the depth axis. In addition, we ignore the refraction of the surface for simplicity. Figure 4.3 shows simulated PSFs by changing the optical thickness T. When T is small (thin scattering media), a sharply pointed PSF is generated, and as T becomes larger (thicker scattering media), the PSF spatially spreads wider.

These PSFs only represent the spread of light but dot not consider the absorption by the medium. When the absorption is homogeneous, it only makes the deeper layers darker and has no effect on the sharpness of the layer recovery except for the siganl-to-noise ratio (SNR), hence it can be practically neglected. However, if the absorption is inhomogeneous, which is the case in inhomogeneous translucent objects, the radiance slice of inner layer exhibits artifacts if the spatial variation of PSFs is ignored. In this work, we explicitly take into account the spatially-varying PSFs in Sec. 4.3.3 for dealing with inhomogeneous translucent objects.

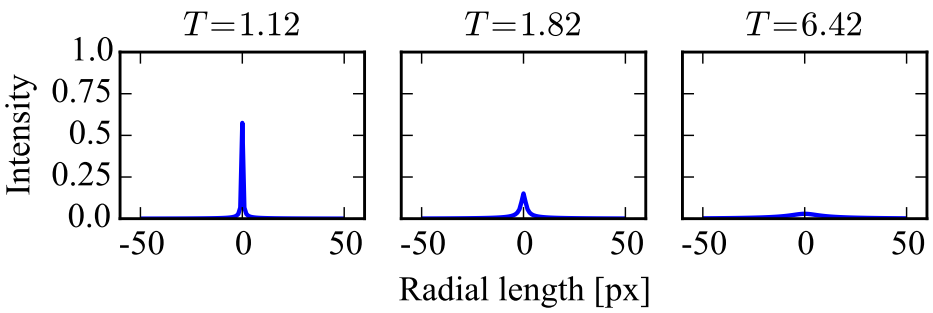


Figure 4.3: Simulated PSF variations using the RTE model with varying optical thickness T

4.3 Proposed Method

We are interested in recovering radiance slices R_d from the mixed observation L_o. To achieve this goal, we develop a *multi-frequency illumination* measurement method, which is built upon the high-frequency illumination (HFI) method proposed by Nayar *et al*. [12]. To begin with, we briefly review the original HFI method.

High-frequency illumination method [12]

The HFI method separates direct and global components by projecting small pitch checker patterns. When the phase of the projection pattern changes slightly, the direct component $D(q)$ varies accordingly, but the global component $G(q)$ remains stable. Based on this observation, their method computes direct and global components using the maximum $L_{\max}(q)$ and minimum $L_{\min}(q)$ intensities that are obtained by shifting the projector pattern as

$$\begin{cases} D(q) = L_{\max}(q) - L_{\min}(q), \\ G(q) = 2L_{\min}(q). \end{cases} \tag{4.8}$$

The direct component $D(q)$ contains high-frequency components, while the global component $G(q)$ contains only lower frequency components than the frequency of projection pattern. Therefore, the HFI method can be viewed as a separation technique for high- and low-frequency components.

Pattern pitch of HFI

In our case, when a translucent object is measured under HFI with pattern pitch p, we can obtain direct component $D(p,q)$ and global component $G(p,q)$ as

$$\begin{cases} D(p,q) &= \sum_d D_d(p,q) \\ G(p,q) &= \sum_d G_d(p,q) \\ S_d(q) &= D_d(p,q) + G_d(p,q), \end{cases} \tag{4.9}$$

where $D_d(p,q)$ and $G_d(p,q)$ are the direct (high-frequency) and global (low-frequency) components at depth d, respectively, and the sum of direct and global components for each depth becomes the radiance slice. As mentioned in [12], the pattern pitch p must be sufficiently smaller than the scene texture for a faithful separation.

We have observed that the direct and global components vary when the projected pattern pitch p changes. Indeed, there is a tight relationship between the pitch size and strength of direct component. Suppose we measure the scene with two distinct pattern pitches p_u and p_v ($p_u < p_v$) independently. When the pattern pitch becomes larger, the wider spread of light rays are included in the direct component, hence the direct components have the following relationship:

$$D_d(p_u,q) < D_d(p_v,q), \quad p_u < p_v. \tag{4.10}$$

This relationship indicates that the separation frequency varies with the pattern pitch – wider frequency band is regarded as the direct component as the pattern pitch becomes larger. As low-pass characteristics of depth-dependent PSFs are also depth-dependent, the difference of direct components $D_d(p_v,q) - D_d(p_u,q)$ also varies with depth d. Our method exploits these differences for recovering the appearance of each inner slice by changing the projection pattern pitch.

4.3.1 Multi-frequency Illumination

By measuring the target object using multi-frequency patterns, multiple of corresponding direct components are obtained. Unfortunately, increasing the number of measurements does not make the problem easier as it also increases the number of variables to solve for. To make the problem tractable, we assume that the texture of direct components does not vary drastically when the pattern frequency is high enough and the pitch variation ($p_v - p_u$) is sufficiently small. These direct components D_d at a certain depth d are supposed to have a similar texture with the original radiance; therefore, we can expect the following relationship:

$$D_d(p,q) \approx \alpha(h_d,p) R_d(q), \tag{4.11}$$

where $\alpha(h_d,p)$ is the relative brightness of $D_d(p,q)$ to $R_d(q)$. We call $\alpha(h_d,p)$ the *direct component ratio* that represents the ratio of direct component's mean intensity to the radiance $R_d(q)$'s mean intensity. Hence, Eq. (4.9) can be rewritten as

$$D(p,q) = \sum_d \alpha(h_d,p) R_d(q). \tag{4.12}$$

These assumptions are based on the fact that the diffuse reflection and subsurface scattering can be regarded as the same physical phenomena [79, 129]; the light scatters on or beneath the surface and eventually bounces off of the material in random directions. Direct components D_d represent total intensities of lights from all the points inside the object whose distance from the incident point on the surface is smaller than the pattern pitch p. Hence the pattern pitch p controls the *scale* of the separation of scattered lights in the direct-global separation scheme, and thus controls the intensity of direct components. Furthermore, because p is sufficiently smaller than the scene texture, the texture of direct components and the original texture are largely similar. Based on these observations, we obtain the original texture at different brightnesses by changing p.

With these assumptions, a set of direct component images $D(p, q)$ taken under the multi-frequency illumination of m pitch variations ($p = p_1, p_2, \ldots, p_m$) can be written in a matrix form as

$$\boldsymbol{D}(q) = \boldsymbol{A}\boldsymbol{R}(q), \tag{4.13}$$

where

$$\begin{aligned}
\boldsymbol{D}(q) &= \begin{bmatrix} D(p_1, q) & D(p_2, q) & \cdots & D(p_m, q) \end{bmatrix}^T, \\
\boldsymbol{A} &= \begin{bmatrix} \alpha(h_{d_1}, p_1) & \cdots & \alpha(h_{d_n}, p_1) \\ \vdots & \ddots & \vdots \\ \alpha(h_{d_1}, p_m) & \cdots & \alpha(h_{d_n}, p_m) \end{bmatrix}, \\
\boldsymbol{R}(q) &= \begin{bmatrix} R_{d_1}(q) & R_{d_2}(q) & \cdots & R_{d_n}(q) \end{bmatrix}^T.
\end{aligned}$$

Here, $\boldsymbol{D} \in \mathbb{R}^m$ is a vector of direct components measured under m variations of the pattern pitches at pixel q, $\boldsymbol{R} \in \mathbb{R}^n$ is a vector of n layers of radiance slices, and $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is a matrix containing direct component ratios computed from the projected pattern pitch and the depth-dependent PSF.

When the number of projected patterns m is no less than the number of depth layers n ($m \geq n$) and $\mathrm{rank}(\boldsymbol{A}) = n$, the radiance slices $\boldsymbol{R}$ can be obtained by a norm approximation of the residual vector, *i.e.*, $\boldsymbol{D}(q) - \boldsymbol{A}\boldsymbol{R}(q)$. For example, with a least-squares approximation, the radiance slices $\boldsymbol{R}(q)$ can be determined using the pseudo-inverse $\boldsymbol{A}^+$ as

$$\boldsymbol{R}(q) = \boldsymbol{A}^+\boldsymbol{D}(q). \tag{4.14}$$

Computation of direct component ratio

The direct component ratio $\alpha(h_d, p)$ can be derived from the depth-dependent PSF h_d and the projected pattern pitch p. When a checker pattern is projected to a translucent object, it reaches the depth d with some blur effect and returns to the surface with the additional blur effect. To obtain the direct component ratio, we consider the difference between maximum and minimum intensities in a similar manner to the original HFI as :

$$\begin{aligned}\alpha(h_d, p) = &\max\left((\Pi_p * h_d) * h_d\right) \\ &- \min\left((\Pi_p * h_d) * h_d\right),\end{aligned} \tag{4.15}$$

where Π_p is normalized projection patterns, whose pitch is p. The normalized illumination is defined in the range between 0 and 1; 0 being black and 1 being white. The operators $\max$ and $\min$ return the maximum and minimum value from all pixels, respectively.

4.3.2 Estimation of Informative Slices

Once we know the target depths (or PSFs) to inspect, we can set up a matrix $\boldsymbol{A}$ using Eq. (4.15), and thus can recover slices corresponding to the depths using Eq. (4.14). However, such a prior knowledge is difficult to obtain before measurement; therefore, automatically selecting a good set of depths becomes important for recovering *informative* slices. For example, if an arbitrary depth is chosen, it has a chance to correspond to the middle of distinct texture layers. To recover informative slices, we use a two-step approach. The first step is the estimation of a set of informative depths via optimization. This is equivalent to selecting a small number of useful PSFs from many other possible PSFs. The second step is the recovery of slices at the informative depths determined by the earlier step. The overall procedure of this strategy is illustrated in Fig. 4.4. Now, we explain the details of each step.

Step 1: Estimation of informative depths

Elements of matrix $\boldsymbol{A}$ depend on depth-dependent PSFs; hence estimation of informative depths corresponds to estimating the shape of PSFs. The candidate relative depth T, which determines the shape of the PSFs, are set in the range between $1 + \epsilon$ and $10 + \epsilon$ with a step size 0.05. We add the offset $\epsilon\ (= 0.01)$ because Eq. (4.6) does not converge if $T \leq 1$. Initially, we set up the matrix $\boldsymbol{A}$ using all candidate parameters T (thus $m < n$) in order to estimate informative slices.

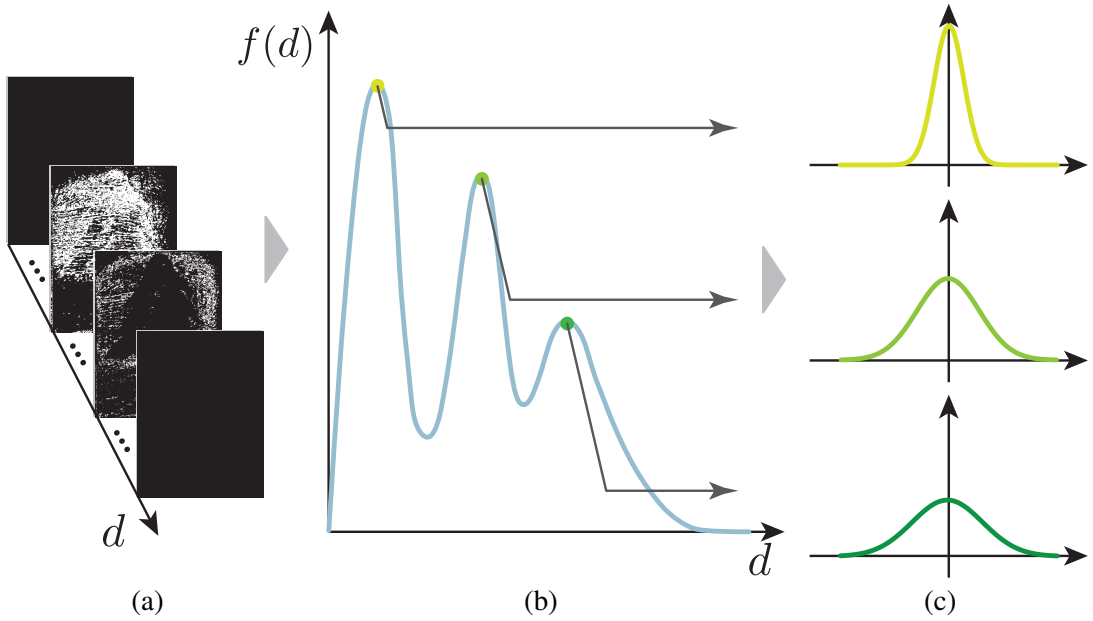


Figure 4.4: Selection of informative PSFs (corresponding to certain depths). (a) Estimates $\hat{R}_d$ obtained via optimization (Eq. (4.16)). Non-zero pixels indicate the informative regions. (b) We find the local maxima of the non-zero pixel counts (corresponding to l_0-norm of $\hat{R}_d$) for all pixel coordinates across depth. (q) Selected PSFs corresponding to the local maximas

Frequently, there are only a small number of informative slices inside translucent objects. Hence we can regard such radiance slices exist sparsely along depth. Our method uses this sparsity to determine the informative slices by solving a l_1 regularized problem (as known as the lasso [130]) with a non-negative constraint about $\boldsymbol{R}$:

$$\hat{\boldsymbol{R}}(q) = \underset{\boldsymbol{R}(q)}{\operatorname{argmin}} \|\boldsymbol{A}\boldsymbol{R}(q) - \boldsymbol{D}(q)\|_2^2 + \lambda \|\boldsymbol{R}(q)\|_1 \quad \text{subject to } \boldsymbol{R} \succeq 0. \tag{4.16}$$

We can regard the depth $d\,(= \frac{T}{\sigma})$, where $\hat{R}_d(q)$ has a non-zero value, is informative while others are not. Equation (4.16) becomes a quadratic programming (QP) problem and thus can be efficiently solved in a polynomial time. We solve the optimization in a per-pixel manner. Solving a similar problem for the entire image at a time instead of computing in a pixel-wise manner is also a viable option; however, we have observed that they do not make much difference because of the following Step 2. Therefore, for efficient parallelization, we choose the per-pixel implementation.

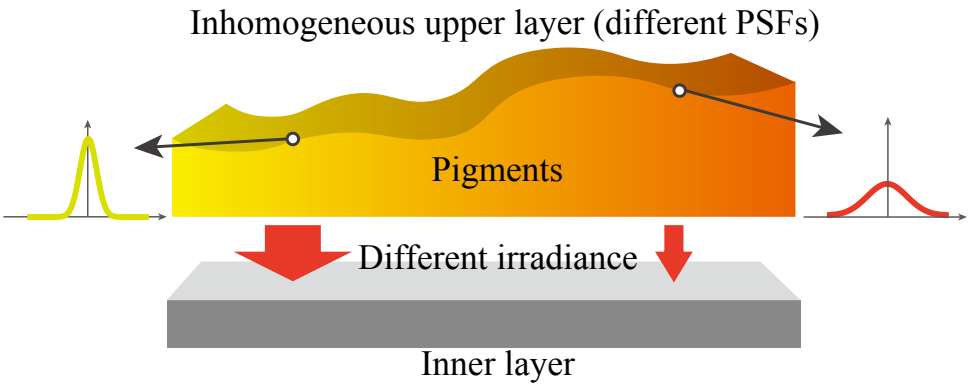


Figure 4.5: Scene of inhomogeneous upper layer. PSFs and irradiance of inner layer vary depending on the upper layer's optical property.

Step 2: Informative slice recovery

This step determines informative depth slices of the whole image by consolidating all the pixel-wise selections.

The informative depths $\hat{\boldsymbol{d}}$ are local maximas of the sum of l_0 norm of $\hat{R}_d(q)$ for all pixels. The evaluation function $f(d)$ is defined as

$$f(d) = \sum_q \|\hat{R}_d(q)\|_0, \tag{4.17}$$

and we find all local maxima of $f(d)$ along d in the rage of interest as shown in Fig. 4.4(b). Once the depths of interest are selected, we can set up a small matrix $\hat{\boldsymbol{A}}$. Finally, the appearances of informative slices are recovered using the matrix $\hat{\boldsymbol{A}}$ using the least-squares approximation as

$$\boldsymbol{R}(q) = \hat{\boldsymbol{A}}^{+}\boldsymbol{D}(q) \tag{4.18}$$

in the same manner to Eq. (4.14). We call the method described so far a *baseline method*.

4.3.3 Recovering Inner Slices inside Inhomogeneous Upper Layer

Now we describe an extension of the baseline method for dealing with inhomogeneous translucent objects. When the upper layer is inhomogeneous, due to either different materials or thicknesses, or even both, recovered inner slices by the baseline method are affected by the upper layer's non-uniform scattering and transmission. Because of the spatially non-uniform scattering and absorption property, both PSF and irradiance of the inner layer vary even if the inner layer is homogeneous as illustrated in Fig. 4.5. Under this situation, the baseline method suffers from artifacts because it assumes a single PSF per slice for the entire image region.

To deal with the inhomogeneous upper layer, we employ a pixel-wise PSF selection approach. By applying Step 2 of the baseline method in a sliding window manner, a set of PSFs can be obtained for each pixel. We apply a 2D Gaussian-weighted filter for the sliding window for suppressing high-frequency observation noise. The evaluation function $f_q(d)$ for pixel location q is defined as

$$f_q(d) = \sum_{q_n \in W_q} \mathcal{N}_{\sigma_f}(\|q_n - q\|_2) \left\| \hat{R}_d(q_n) \right\|_0, \tag{4.19}$$

where $\mathcal{N}_{\sigma_f}$ is a zero-mean Gaussian distribution with standard deviation σ_f centered at q, and q_n is a pixel location in the window W_q. By selecting the informative depths for each pixel via finding the local maxima of f_q, we can recover pixel-wise informative slices in a similar manner to Eq. (4.18) as

$$\boldsymbol{R}(q) = \hat{\boldsymbol{A}}^{+}(q)\boldsymbol{D}(q), \tag{4.20}$$

where $\hat{\boldsymbol{A}}^{+}(q)$ is the pseudo-inverse of a small matrix containing direct component ratios computed for pixel q using the selected PSFs.

The recovered inner slices may still contain the brightness discontinuity due to the non-uniform transmission, because PSFs considered so far neglect absorption by the medium. Inhomogeneous transmittance causes inhomogeneous irradiance on the inner layer, hence the recovered inner layer may have brightness variations. Indeed, this brightness discontinuity stands out at the boarder of distinct PSFs. In this work, we adopt a gradient-based image repairing approach to reduce the brightness discontinuity. Specifically, we suppress the gradient of edges at the brightness discontinuity in the inner with retaining other edges so that a cleaner appearance of the inner slice can be obtained. Such edges $\mathcal{E}$ can be identified by observing the change of PSF estimates because the discontinuity of transmission occurs at the boarder of different materials or thicknesses in the upper layer. Hence, the brightness corrected slice R'_d can be recovered by setting the gradient of such edges to zero as

$$\nabla R'_d(q) = \begin{cases} \mathbf{0} & \text{if } q \in \mathcal{E} \\ \nabla R_d(q) & \text{otherwise,} \end{cases} \tag{4.21}$$

where ∇ is the gradient operator and $\mathcal{E}$ is the set of edges corresponding to the material (or thickness) borders. The integration is computed by a standard sparse linear system solver as done in the work of Poisson image editing [131].

4.4 Experiments

We evaluate our method numerically by simulation and also show qualitative results using real-world scenes.

4.4.1 Evaluation by Simulation

We first assess the appropriateness of the approximation in Eq. (4.11), and then evaluate the accuracy of the slice recovery.

Validity of approximation in Eq. (4.11)

In this simulation, we change the pattern pitch p from 3 to 20 pixels, the depth d from 4 to 18 [mm] (in optical depths T, which correspond to 4σ to 18σ). The scene is one-layered for the purpose of assessing the approximation in Eq. (4.11). The coefficient σ is set to 0.27 [mm^{-1}], and forward scattering parameter g is set to 0.9. These parameters are chosen according to [5]. We generate the appearance slices $S^{\varphi}_{d,p}$ using checker pattern Π^{φ}_{p} as

$$S^{\varphi}_{d,p} = ((\Pi^{\varphi}_{p} * h_d) \circ R_d) * h_d. \tag{4.22}$$

p and φ are the pitch and phase of the pattern, and $\circ$ is the Hadamard (element-wise) product operator. We compute direct component $D_d(p)$ from these synthetic images with changing the phase φ.

We assess correlations between the recovered direct components $D_d(p)$ and the ground truth slices R_d using zero-mean cross correlation (ZNCC). The ZNCC value falls in the range from -1 to 1, where a negative value indicates negative correlation, and a greater value indicates higher similarity. The evaluation results are summarized in Fig. 4.6. The plots correspond to the average of ZNCC scores obtained from 15 different textures R_d. ZNCC values decreases as the depth becomes deeper (greater T) and pattern pitch becomes larger. However, they are all consistently highly correlated (minimum ZNCC value is 0.982 and the mean value is 0.992), and it shows the accuracy of the approximation.

Evaluation of slice recovery

Here we evaluate the overall accuracy of our method via simulation. We generate images L^{φ}_{p} under projection patterns Π^{φ}_{p} as

$$L^{\varphi}_{p} = \sum_{d} ((\Pi^{\varphi}_{p} * h_d) \circ R_d) * h_d. \tag{4.23}$$

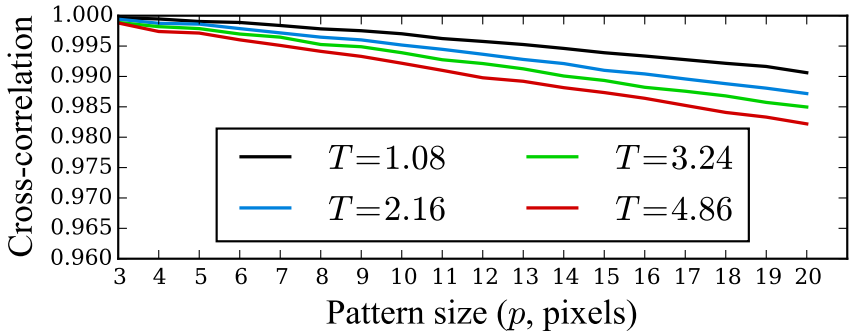


Figure 4.6: Evaluation of the approximation in Eq. (4.11). We calculate the cross correlation between direct component slice $D_d(p)$ and the ground truth slice R_d. It shows high correlation even for the worst case.

Layers	ZNCC values Max.	Mean	Min.
Top layer	0.99	0.93	0.89
Bottom layer	0.92	0.84	0.67

Table 4.1: Two layers recovery result for 20 sets of scenes. We compare recovered slices with the ground truth slices by ZNCC. Higher ZNCC scores indicate more accurate recovery.

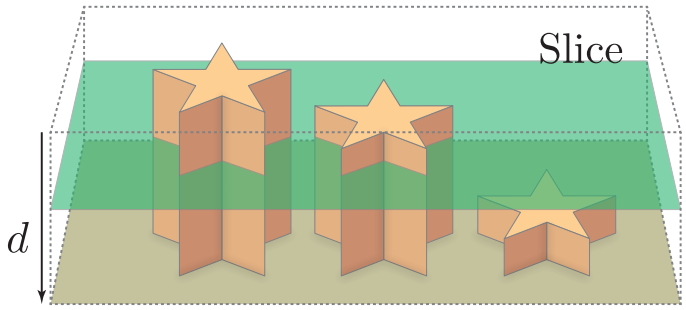
We use 20 different two-layered scenes that have distinct textures R_5 and R_{15} at two depth ranges $5 \leq d < 15$ and $15 \leq d$. The textures are randomly paired from the ones that are used in the previous experiment. We change the pitch p of checker pattern and shift the pattern with φ. From the generated images, we compute $D(p, q)$ for each pattern pitch p, and apply our method to recover slices at $d = 5$ and 15.

We again use ZNCC values between the recovered and ground truth slices. The experimental results are summarized in Table 4.1. In all data sets, the ZNCC values of upper layer is higher than the lower layer as expected. The result of this simulation indicates that our method can recover slices of various textures with high accuracy.

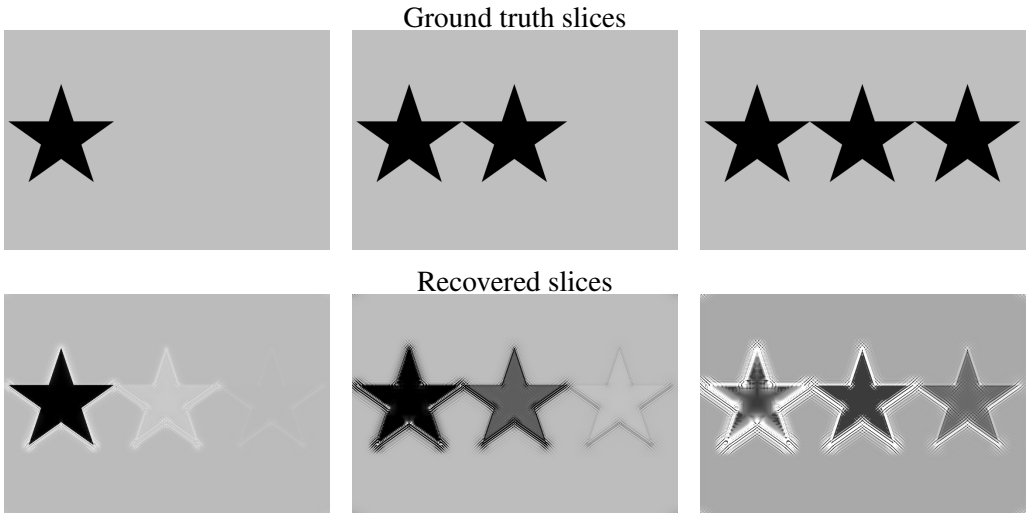
Figure 4.7 shows a synthetic example of three-layer recovery, where slices at at $d = 1$, 5, and 15 are recovered. The ZNCC scores for the recovery results are 0.98, 0.83, and 0.51, respectively. While the result is generally consistent with the ground truth, negative intensities and ringing artifacts are observed due to the discrete pitch variations and convolution.

4.4.2 Real-world Experiment

We develop a coaxial projector-camera setup for realizing the measurement setup as shown in Fig. 4.8. The coaxial setup has a favorable property; correspondence between projector



(a) Simulated scene. three star-shaped pillars with different heights are placed in scattering media. An example slice is illustrated by the green plane.



(b) Comparison of the recovered and ground truth slices

Figure 4.7: (a) Target scene (b) Result of three-layer recovery at $d = 1$, 5, and 15. The ground truth radiances slices (upper) and recovered slices (lower) are shown. ZNCC scores are 0.98, 0.83, and 0.51, respectively.

and camera pixels becomes invariant with respect to depths. Unlike non-coaxial settings, with which a illumination ray inside the translucent object forms a line in the image coordinates [16], the coaxial setting allows us to easily separate the direct rays. We use a LightCommander projector, which is a DMD projector development kit by Texas Instruments, and use near infrared (NIR) light for measurements. The lenses of both camera and projector are set equivalent (Ai Micro-Nikkor 105mm f/2.8S) for making the alignment easy. In the experiment, we use 18 variations of checker patterns (3px to 20px with 1px interval), and shift the pattern for one-third of square size for each pattern.

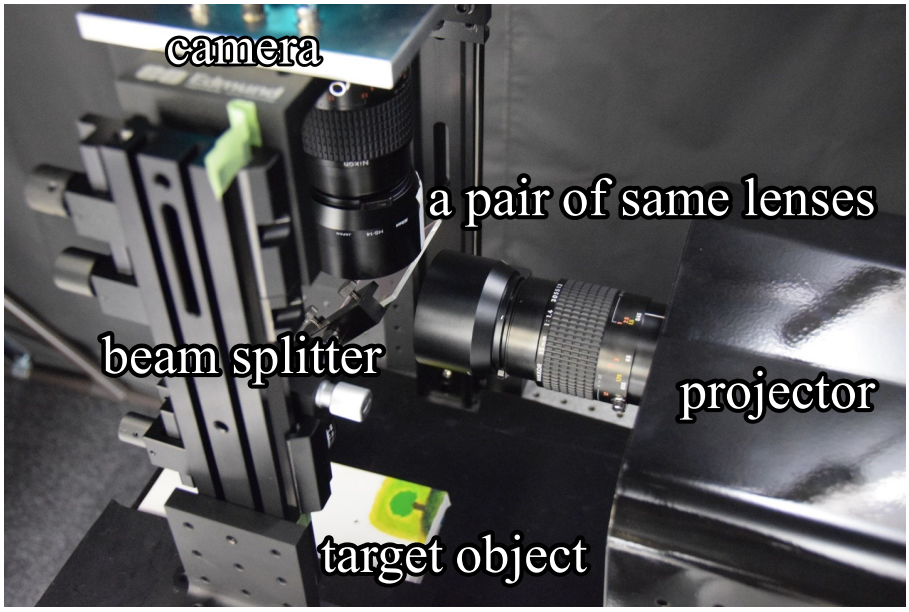
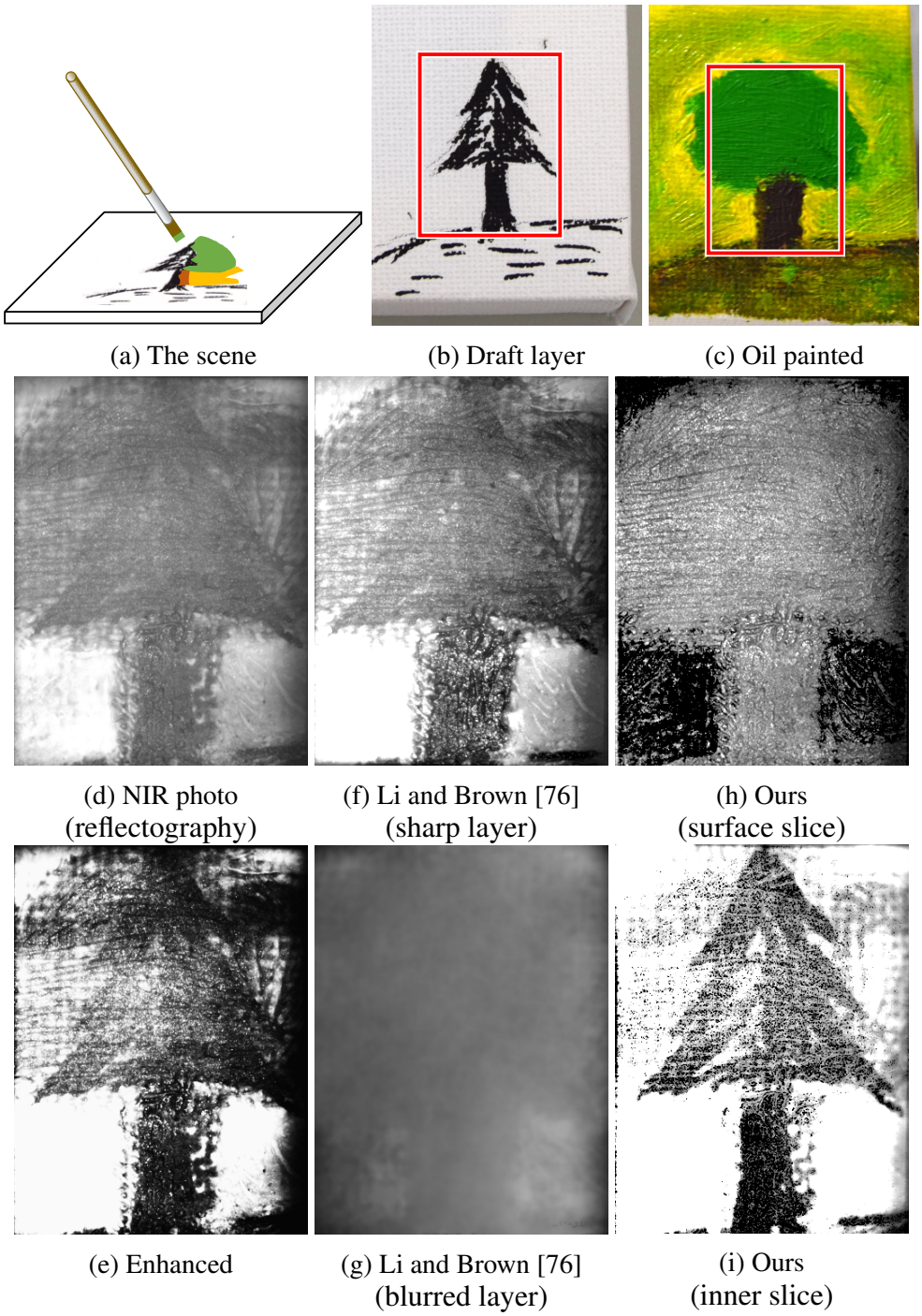


Figure 4.8: Measurement setup. The coaxial system allows us to maintain the correspondences between projector and camera pixels.

Experimental results

First, we use an oil painting as a target scene as shown in Fig. 4.9(a), which has draft and surface pigment layers as depicted in Figs. 4.9(b) and 4.9(c). By taking a standard photograph under the near infrared light (as done in infrared reflectography [84] in the art field), we can only vaguely observe the draft layer as shown in Fig. 4.9(d). Since it is the superposition of draft and surface pigment layers, it naturally results in a blurry image. Even with a manual contrast adjustment, it is difficult to clearly observe the shape of draft tree as depicted in Fig. 4.9(e). A simple layer decomposition still suffers from the blur as shown in Figs. 4.9(f) and 4.9(g). By applying our method to this scene, two PSFs are estimated as depicted in Fig. 4.10(a) and two slices are recovered as in Figs. 4.9(h) and 4.9(i). The upper surface layer corresponds to the surface pigment layer. Because the yellow pigment is almost transparent in the infrared wavelength, the corresponding painting regions become dark in the surface slice. The lower layer shows the inner layer, where the texture of the tree is clearly observed. The variation of the remaining high-frequency components of each layer, which can be computed from its PSF using Eq. (4.15), along with the pattern pitch is shown in Fig. 4.10(b). It shows that the high-frequency components of the upper layer remain more significantly than that of the inner layer, and it indicates the separability of the layers.

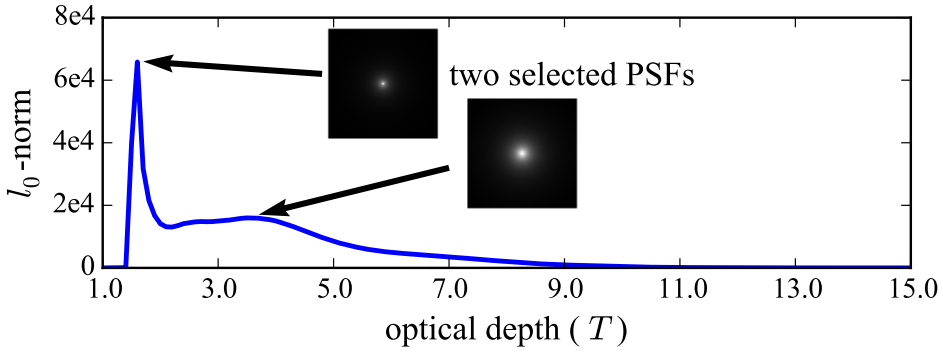


(a) The scene (b) Draft layer (c) Oil painted

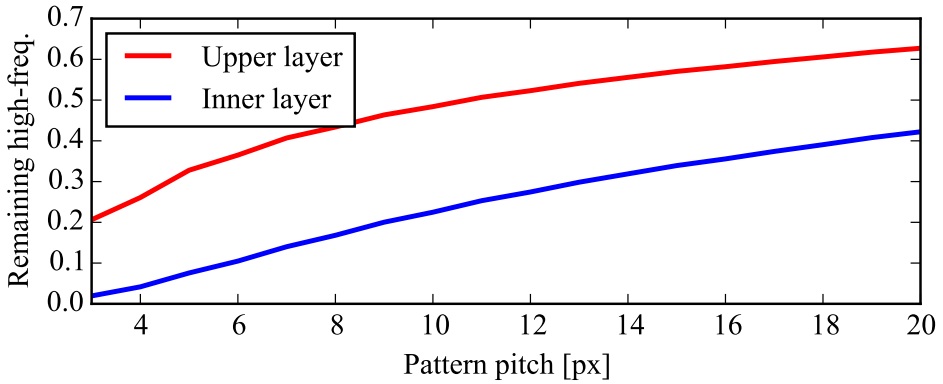
(d) NIR photo (reflectography) (f) Li and Brown [76] (sharp layer) (h) Ours (surface slice)

(e) Enhanced (g) Li and Brown [76] (blurred layer) (i) Ours (inner slice)

Figure 4.9: Experimental result of oil painting using the baseline method. (a) Target scene. We draw a colored round tree on top of the draft of spiny tree. (b) Inner layer (draft) of the painting. (q) Painted scene. Red rectangle region is measured. (d) Normal photo using infrared light. (e) Enhanced image. Intensity range and contrast of (d) are manually adjusted. (f, g) Layer separation results of Li and Brown [76]. Because their method separates sharp and blurred layers, it suffers from global components. (h, i) Results of our method. Layer of surface texture and hidden drawing, respectively. Range of the intensities is adjusted for visualization.



(a) Selected PSFs corresponding to local maxima

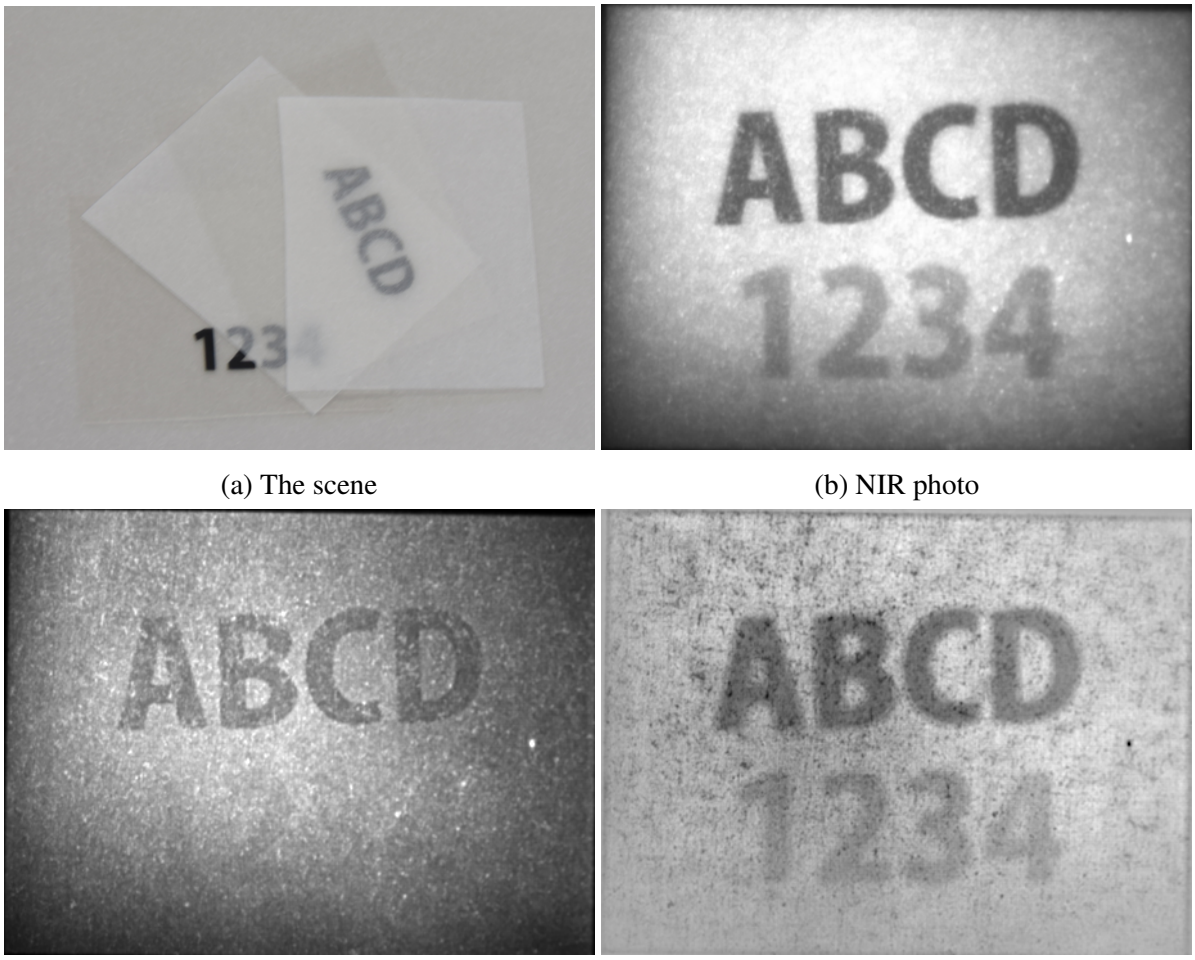


(b) Remaining high-frequency components for each layer

Figure 4.10: Selected PSFs and remaining high-frequency components for each layer. (a) Selected PSFs. There are two peaks in the plot of Eq. (4.17), hence two corresponding PSFs are selected to recover. (b) Remaining high-frequency components for each layer. Upper layer remains more high-frequency components than that of the inner layer.

The second target scene is stacked translucent sheets that consist of layered translucent sheets and printed texture films as shown in Fig. 4.11(a). With a conventional NIR photograph, we can observe mixed textures as shown in Fig. 4.11(b), where textures are blurry and appearances across depths are superposed. Our method correctly selects two informative depths and recovers their radiance slices as shown in Figs. 4.11(c) and 4.11(d). In the upper slice, only 'ABCD' texture is visible, and '1234' texture appears in the lower slice. Due to the shadowing effect caused by the opaque material on the upper layer (the texture of 'ABCD'), the lower slice contains the 'ABCD' texture as an unobserved shadowed region.

Additional results are shown in Fig. 4.12. The top row is a piece of painting, where the painter's signature is hidden under pigment. Our method clearly recovers surface texture and inner signature slices. The middle row is a mural painting covered by white mold. In this example, we used RGB light sources instead of NIR, and record each color channel separately.



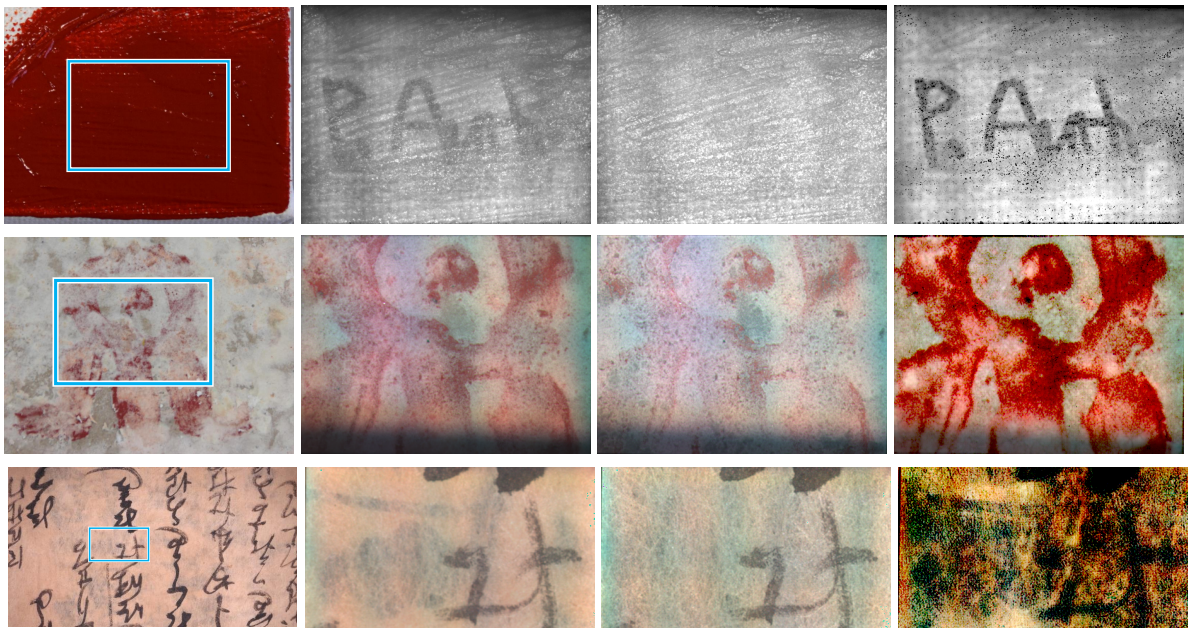
(a) The scene (b) NIR photo

(c) Upper slice (d) Lower slice

Figure 4.11: Experimental result of layered scene using the baseline method. (a) Scene is a composite object of texture and translucent sheets (tracing paper). (b) Normal photo of the scene. Textures inside the object can be seen, although blurry. (q) Recovered slice of upper layer. (d) Recovered slice of lower layer. Textures in sub-millimeter gap can be separated. Further analysis of shadowed region is shown in Fig. 4.14

Our method can recover the slices in this example as well, and the inner appearance is clearly visible in the result of Fig. 4.12(d). The bottom row is a double-sided ancient document with RGB light sources. While an ordinary digital scanner suffers from the back-side texture, our method can faithfully separate surface and back-side textures.

Figure 4.13 shows the result of the scene with an inhomogeneous upper layer. We use a 21×21 pixels Gaussian window for selecting PSFs. The target object is an oil painting, where the upper layer consists of a wide variety of pigments in their thicknesses and materials. The result of the baseline method shown in Fig. 4.13(b) exhibits brightness discontinuity at the middle of image region due to inhomogeneity of the upper pigments. By selecting PSFs



(a) Target scene (b) Normal photo (c) Surface slice (d) Inner slice

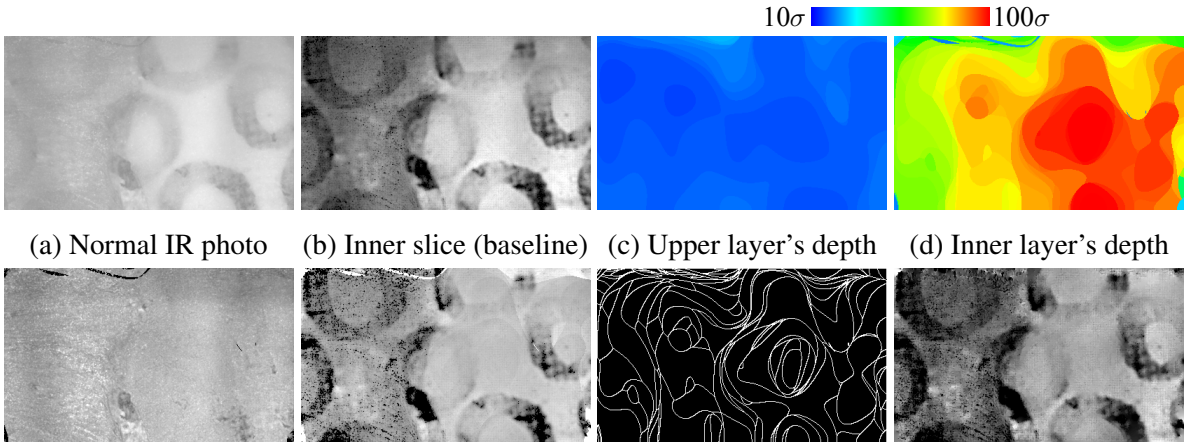
Figure 4.12: Additional results for a painting, a mural, and an ancient document using the baseline method. (a) Target scenes. Top: author's signature covered by red pigment. Middle: A mural painting covered by white mold. Bottom: An ancient document of double side print. Rectangle regions are measured. (b) Normal photograph. Both upper and lower slices, and global components are composed. (q) Recovered slices of surface. (d) Recovered inner slices.

per pixel using the Gaussian-window, the inner layer is recovered in a pixel-wise manner as shown in Fig. 4.13(f), although it still suffers from brightness discontinuity. Finally, with the gradient-domain filtering to suppress the brightness discontinuity, the inner layer that has uniform brightness can be recovered as shown in Fig. 4.13(h). It shows improvement in the visual quality, especially in the left part of the image and overall contrast.

Shadow detection from recovered inner layers

The recovered appearance of inner layers suffers from shadows caused by their upper layers. If the upper layer's transmittance is low, it casts shadow to the inner layer, *e.g.*, as 'ABCD' scene shown in Fig. 4.11(d). Here we describe a simple technique for dealing with shadows by post-processing. To simplify the discussion, we assume that there are only two layers in the scene, *i.e.*, upper and inner layers.

The shadowed regions have two common properties: (1) The shape of shadows observed in the inner slice is similar to the texture of upper slice. (2) The shadowed region becomes darker due to low irradiance. Based on these observations, we define a shadow likelihood measure P_s for identifying shadowed regions.



(a) Normal IR photo (b) Inner slice (baseline) (c) Upper layer's depth (d) Inner layer's depth

(e) Upper slice (pixel-wise) (f) Inner slice (pixel-wise) (g) Gradient mask (h) Inner slice (post-processed)

Figure 4.13: Result of a scene with an inhomogeneous upper layer using the extended method. (a) Captured image using near infrared lights. (b) Recovered inner slice by the baseline method. (c, d) Selected depths for upper and inner layers, respectively. (e) Recovered upper slice by pixel-wise recovery. (f) Recovered inner slice by pixel-wise recovery. (g) Gradient mask computed from (d) that is used for gradient-domain filtering. (h) Recovered inner slice from (f) and (g) by image enhancement in the gradient domain.

The similarity S of the texture shapes in the recovered slices R_1 and R_2 can be obtained using the absolute value of their cross correlation as

$$S(q) = |C_{w,c}(R_1, R_2)|\,, \tag{4.24}$$

where $C_{w,c}$ is the cross-correlation within a small window w centered at pixel location q, and values in R_1 and R_2 are normalized in the range of $[0, 1]$. Using this similarity, the shadow likelihood P_s is defined using the darkness of the shadowed region in R_2 as

$$P_s(q) = S(q)(1 - R_2(q)). \tag{4.25}$$

It yields a likelihood score for each pixel being in shadow, and by a simple thresholding, a shadowed region can be determined.

Once the shadow regions are identified, we can use a shadow removal technique. For example, the lost information within the shadow region can be filled in by arbitrary image inpainting methods. In this paper, we use a patch-match based image inpainting method [132].

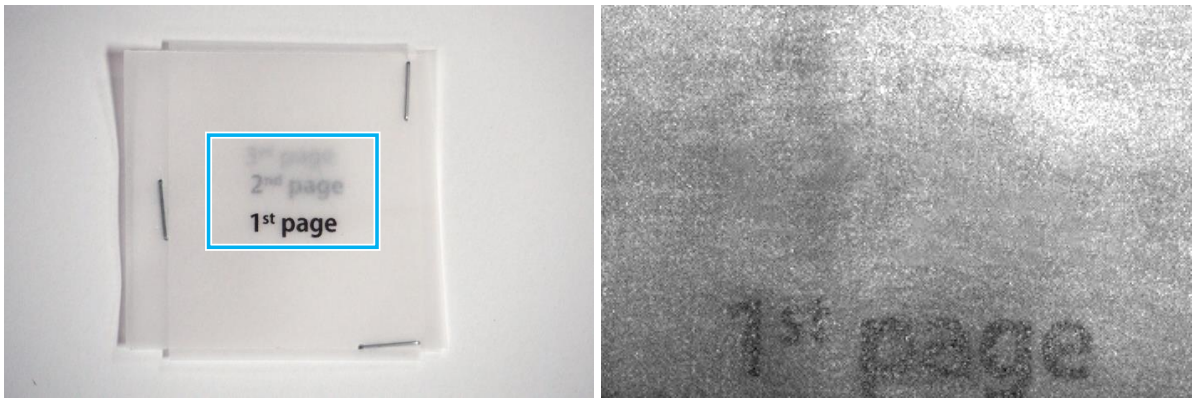
The result of shadow detection and removal for the 'ABCD' and ancient document scenes are shown in Fig. 4.14. For this experiment, we use 21×21 window size for computing the similarity measure. The shadow likelihood shown in Fig. 4.14(b) qualitatively corresponds

(a) Inner slice including shadow (b) Likelihood of shadow (c) Shadow region mask (d) Inpainted inner layer

Figure 4.14: Shadow detection and inpainting results. The scene are the same as Fig. 4.11 and Fig. 4.12. (a) The inner layer including shadowed region. (b) Shadow likelihood from recovered slices. White pixels indicate higher likelihood of being in shadow. (q) Shadow region is indicated by red by thresholding the likelihood. (d) Shadow region is recovered by image inpainting [132]. Shadow artifact caused by the upper layer is suppressed.

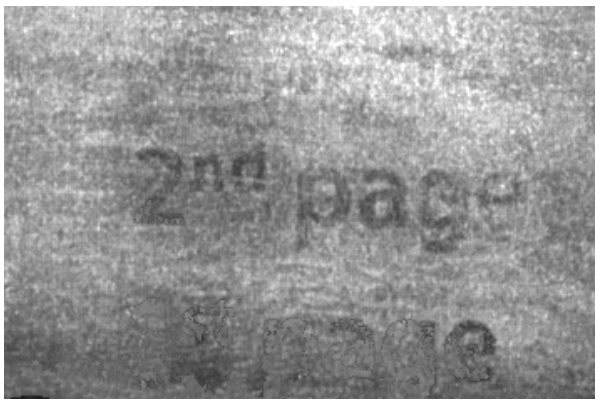
to the shadowed regions in the inner layer. The shadow mask generated by thresholding the shadow likelihood is shown in Fig. 4.14(c). Figure 4.14(d) shows the image inpainting result using the detected shadow mask and the method of [132]. It shows improvement of visual quality of the inner layer.

Finally, we show a recovery and shadow removal result for a three-layer scene. Figure 4.15(a) shows the target scene, where three printed papers are overlaid. By applying our baseline method, three slices with shadows are recovered. With shadow detection and inpainting, we obtain the recovery of three layers as shown in Figs. 4.15(b)-4.15(d) For the shadow detection of the third slice, we compute the correlation with the original second slice which contains the shadows casted by the top slice. Figure 4.15(e) shows how high-frequency components remain along with varying pattern pitches. The lowest layer only retains a limited amount of high-frequency components hence its recovery becomes slightly blurry. However, it shows the different characteristics among layers, which indicates the separability of the layers.

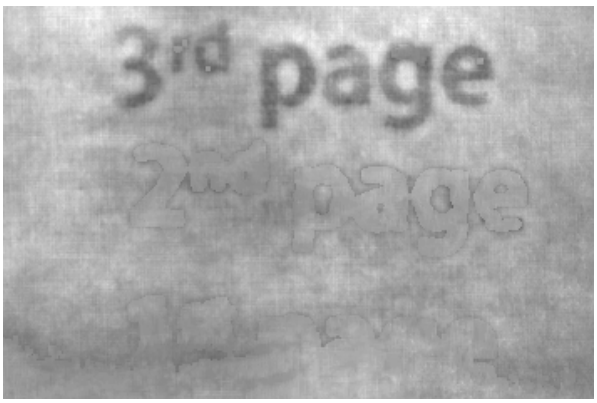


(a) The scene

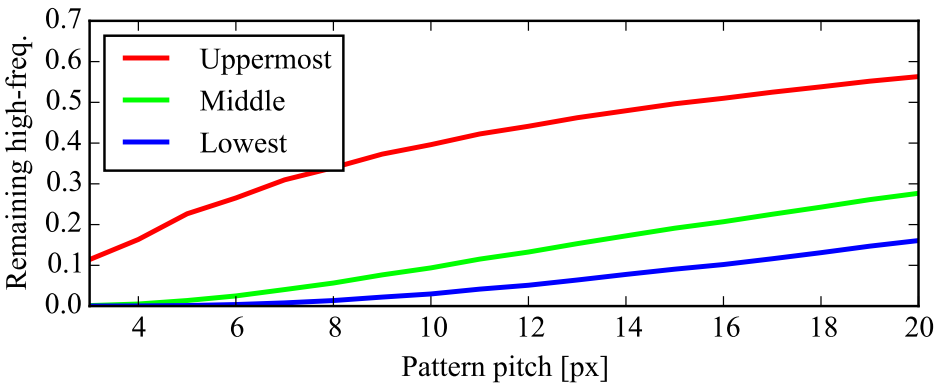
(b) Uppermost slice



(c) Middle slice



(d) Lowest slice



(e) Remaining high-frequency components

Figure 4.15: The result of seeing through pages. (a) The target object. Three printed thin papers are superposed. (b) Recovered uppermost slice. Only the first page can be seen. (q) Recovered second page. Shadows from the upper layer is inpainted. (d) Recovered third page. Shadows from the first and second layers are inpainted. (e) Remaining high-freqency components for each pixel.

4.5 Discussions

This chapter described a method for recovering inner slices of translucent objects based on multi-frequency pattern projection. The proposed method is built upon the observation that the PSF inside translucent objects varies according to the depth of slices. Based on that, we have shown that inner radiance slices can be recovered by estimating PSFs using varying pitches of projection patterns. We also developed a method for automatically selecting informative slices via a sparse representation, *i.e.*, determining sparse coefficients that corresponds to radiance slices. We further extended the method for dealing with inhomogeneous translucent objects based on a combination of pixel-wise appearance recovery and gradient-based image repairing. The effectiveness of the proposed method are shown by several experiments on simulation and real-world translucent objects.

The proposed method is based on a difference of PSF among layers, hence eligible target objects are limited by optical thickness so that every layer has different PSF. Inner layers inside both optically and physically thin translucent objects cannot be recovered because different PSFs are hardly produced, and both thick objects are also ineligible because inner layer is not observable. Related to this problem, the number of recoverable layers is also limited in practice because of PSF variations. Selecting wavelength of the light source, where the target object has appropriate translucency, is one of the approaches to generate distinguishable PSFs among layers.

Chapter 5

Light Path Analysis of Time-of-Flight Distortion

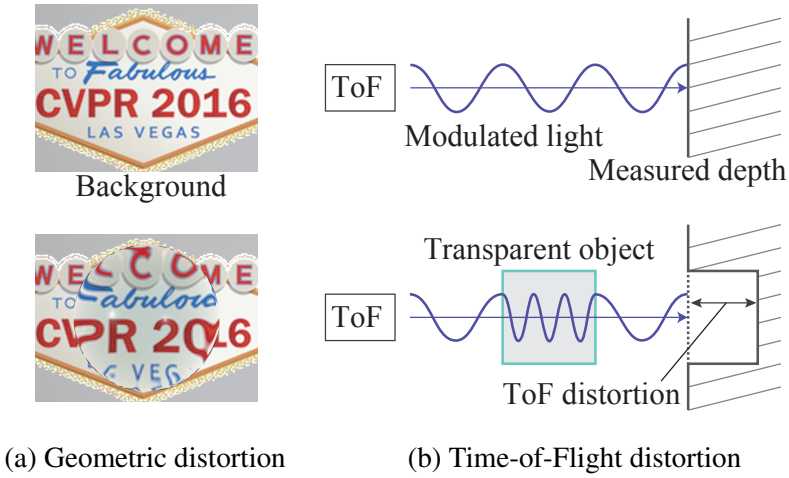
In this chapter, we focus on the temporal spread of light and temporally amplitude-modulated light. We use a Time-of-Flight camera to analyze the scene temporally, which emits temporally modulated light. In this chapter, we propose a method for recovering 3-d shape of a transparent object such as glasses. Because the speed of light slows down according to refractive index, the light is observed later, which distort the Time-of-Flight measurement. Using this temporal delay as a cue for the object's thickness, the 3-d can be recovered.

5.1 Preliminary

Transparent shape reconstruction is important for scientific imaging and applications in industrial manufacturing. It has been a difficult problem in computer vision because the appearance of a transparent object can only be indirectly observed by the distortion of background textures as illustrated in Fig. 5.1(a). Several methods that uses the observations of the *geometric distortion* have been proposed, yet it is still an active research subject.

Recently, a Time-of-Flight (ToF) camera, which measures distance by correlation of its modulated light, is becoming a commodity device. When a scene containing a transparent object is recorded by a ToF camera, the ToF measurement also becomes distorted because the light slows down inside the transparent object due to its refractive index as illustrated in Fig. 5.1(b). The distortion is different than the geometric distortion on the image coordinates, but still conveying rich information about the shape of transparent object. We call this distortion a *Time-of-Flight distortion*[1] in this chapter and use it for recovering the shape of transparent object.

[1]This is not ordinary depth distortions that are due to calibration or multi-path effects but the distortion of "time-of-flight."



(a) Geometric distortion (b) Time-of-Flight distortion

Figure 5.1: Two distortions caused by transparent objects. (a) Background texture is geometrically distorted by refraction. (b) Measured depth is distorted by slower light speed in the medium.

Our method records two ToF measurements of a transparent object from a single viewpoint but by moving the background reference surface which is calibrated. Assuming that the refractive index of the target object is known and the ToF camera is calibrated, the shape estimation problem can be viewed as the problem of searching the light path, which correspond to estimating front and back refraction points and its surface normal. We show that, using the ToF distortion, the light path has a simple expression that is governed by a single parameter. We develop a method for estimating this parameter using a *surface normal consistency*, that represents a consistency between the surface normal computed from the light path candidate and that obtained from the corresponding shape.

The proposed method estimates both front and back surfaces in a single viewpoint approach. Unlike previous single viewpoint approaches that are restricted to a scene with a single refraction, or requiring a number of light sources to illuminate the scene, the proposed method is able to recover a scene with two refraction surfaces from a single view point, with two observations obtained by moving the background reference surface. This new setting is enabled by the use of ToF distortion, which explicitly encodes the altered light speed in the transparent medium and its volume. Furthermore, we show a simple multi-path mitigation technique using a retroreflective sheet for this setting, which does not require any computational illumination devices, for recovering curved or multi-planar surfaces of transparent objects.

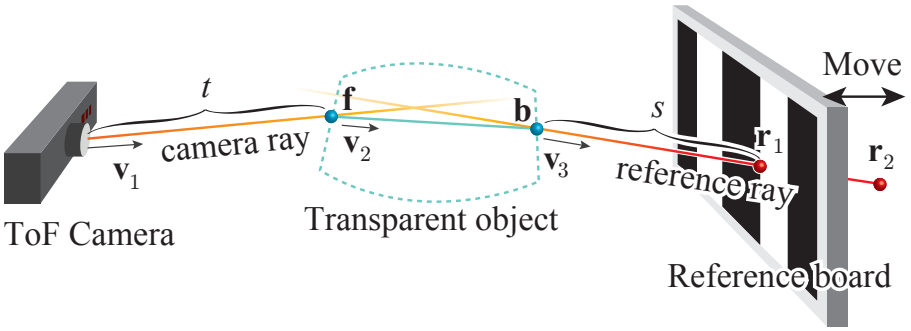


Figure 5.2: Transparent shape recovery problem. The front surface point $\mathbf{f}$ is on the camera ray at distance t, and the back surface point $\mathbf{b}$ is on the ray of 2 reference points at distance s.

5.2 Proposed Method

In our setting, we measure a transparent object using a known reference board placed behind the object as illustrated in Fig. 5.2. The target scene is recorded by a ToF camera twice by moving the reference board at two distinct locations behind the object. A single ToF observation contains both intensity and depth measurements, and from the intensity measurement, we determine two reference points $\mathbf{r}_1, \mathbf{r}_2 \in \mathbb{R}^3$. From these two reference point observations, the reference ray direction $\mathbf{v}_3$ can be determined. With an assumption that the refractive index ν of the transparent object is known and that the camera is calibrated therefore the camera ray direction $\mathbf{v}_1$ is known, our goal is to estimate the front surface point $\mathbf{f}$ and back surface point $\mathbf{b}$ at every camera pixel using the ToF depth measurement $\tilde{d}$. This problem is equivalent to estimating two variable t and s, where t is the distance from the camera to the front surface point, and s is the distance from the back surface point to the reference point. Our method estimates these unknowns t and s from $\mathbf{v}_1$, $\mathbf{v}_3$, $\mathbf{r}_1$, and $\tilde{d}$ for determining the shape of transparent object. We begin with describing the ToF distortion model and develop the estimation method.

5.2.1 Time-of-Flight Distortion Model

A ToF sensor acquires the scene depth by observing the delay of returned light. When a transparent object is placed in the scene, a ToF sensor measures the optical length to the background object instead of the transparent object because light refracts and passes through the transparent object and reflects back from the background. In addition, the light speed slows down inside the transparent object according to its refractive index ν, hence the optical length measured by the ToF camera becomes different than the geometric length of the light path. The measured depth $\tilde{d}$ can, therefore, be expressed as

$$\tilde{d} = t + \nu \left|\mathbf{b} - \mathbf{f}\right| + s, \tag{5.1}$$

where $|\cdot|$ represents the geometric length of the vector. We call this distortion the *Time-of-Flight distortion* in this chapter, and use it for estimating the shape of transparent objects as it embeds the thickness of transparent object.

This model is built upon the following two assumptions.

- Reflection on the transparent object surface is ignored. Specular reflection is only observed when the surface confronts with the ToF camera, which rarely happens in practice.
- We assume only a single path of ToF light rays and ignore multi-path interference. In practice, we avoid the multi-path effect using a retroreflective sheet, which is explained in an experiment section. Prior methods for multi-path separation in ToF sensing such as [40, 39] can be alternatively used.

5.2.2 Baseline Method

By examining the equation of the ToF distortion (Eq. (5.1)) analytically, s can be expressed as a function of t as[2]

$$s(t) = \frac{-h(t) - \sqrt{h^2(t) - gi(t)}}{g}, \tag{5.2}$$

where g, h, and i are auxiliary variables defined as

$$\begin{cases} g & = \nu^2 - 1 \\ h(t) & = \tilde{d} - t - \nu^2(\mathbf{r}_1 - t\mathbf{v}_1)^T\mathbf{v}_3 \\ i(t) & = \nu^2 \left|\mathbf{r}_1 - t\mathbf{v}_1\right|^2 - (\tilde{d} - t)^2. \end{cases} \tag{5.3}$$

While this expression eliminates unknown vectors $\mathbf{b}$ and $\mathbf{f}$, there still remains one degree of freedom to determine the unique shape s and t. For resolving the ambiguity, we use a surface normal consistency described in the following.

When a hypothesized depth t is assumed, we can obtain a hypothesized front surface point $\mathbf{f}$ and back surface point $\mathbf{b}$ as functions of t:

$$\begin{cases} \mathbf{f}(t) & = t\mathbf{v}_1, \\ \mathbf{b}(t) & = \mathbf{r}_1 - s(t)\mathbf{v}_3. \end{cases} \tag{5.4}$$

Based on this, the refractive ray direction $\mathbf{v}_2$ can be obtained from the hypothesized front surface point $\mathbf{f}$ to back surface point $\mathbf{b}$ as

$$\mathbf{v}_2(t) = \frac{\mathbf{b}(t) - \mathbf{f}(t)}{|\mathbf{b}(t) - \mathbf{f}(t)|}. \tag{5.5}$$

[2]Please refer to Appendix B.1 for the derivation.

Since the refractive index ν is known, the surface normal $\mathbf{n_p}(t)$ of the front surface point can be obtained from the refractive path using Snell's law as[3]

$$\begin{aligned}\mathbf{n_p}(t) &= \frac{\nu\mathbf{v}_2(t) - \mathbf{v}_1}{|\nu\mathbf{v}_2(t) - \mathbf{v}_1|} \\ &= \frac{1}{N_2(t)}\left(\frac{\nu}{N_1(t)}\left(\mathbf{r}_1 - s(t)\mathbf{v}_3 - t\mathbf{v}_1\right) - \mathbf{v}_1\right),\end{aligned} \tag{5.6}$$

where $N_1(t) = |\mathbf{b}(t) - \mathbf{f}(t)|$ and $N_2(t) = |\nu\mathbf{v}_2(t) - \mathbf{v}_1|$ are normalization coefficients.

At the same time, we can obtain another surface normal $\mathbf{n_d}(t)$ from the hypothesized shape as

$$\mathbf{n_d}(t) = \frac{\frac{\partial(t\mathbf{v}_1)}{\partial x} \times \frac{\partial(t\mathbf{v}_1)}{\partial y}}{\left|\frac{\partial(t\mathbf{v}_1)}{\partial x} \times \frac{\partial(t\mathbf{v}_1)}{\partial y}\right|}, \tag{5.7}$$

where $\times$ and ∂ are cross product and partial differentiation operators, respectively. We assume that the neighbors around $t\mathbf{v}_1$ correspond to those viewed in the camera pixel coordinates.

If the assumed front depth t is correct, two normals $\mathbf{n_p}(t)$ and $\mathbf{n_d}(t)$ should coincide; therefore, the estimation problem can be casted as an optimization problem as

$$\underset{\mathbf{t}}{\operatorname{argmin}} \sum_{c \in C} \|\mathbf{n}_{\mathbf{p},q}(t_c) - \mathbf{n}_{\mathbf{d},q}(t_c)\|_2^2, \tag{5.8}$$

where $\mathbf{t}$ is a vector listing t_c for all pixels, C is a set of all pixels, t_c is the hypothesized front depth of pixel c, $\mathbf{n}_{\mathbf{p},q}$ and $\mathbf{n}_{\mathbf{d},q}$ are the surface normal computed from the refractive path and that from hypothesized shape at pixel c, respectively. We call this method a *baseline method*, and in the next subsection, we introduce additional objectives for a more stable solution.

5.2.3 Robust Estimation Method

The objective function (Eq. (5.8)) directly uses the optical length measured by the ToF sensor (Eq. (5.1)) without considering uncertainty (or noise) in the measurement. To take into account observation noise, we introduce a new variable l_c for each pixel c, which represents the noise-free ToF optical length. There are two objective terms; one of which is the normal consistency used for the baseline method described above, and the other is that the denoised signal l_c should remain close enough to the measured signal $\tilde{d}$. We additionally regularize the denoising part assuming the smoothness of both front and back surfaces. Hence, the overall objective function

[3]Please refer to Appendix B.2 for computing normal from refractive path.

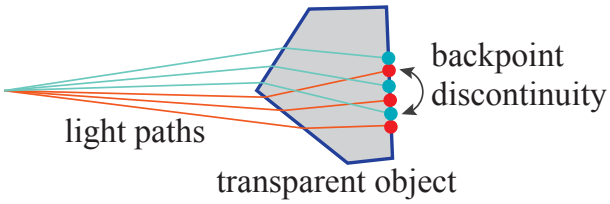


Figure 5.3: Discontinuity of back surface points. Red and blue points are well ordered in each group, but generates discontinuity due to the edge on the front surface as emphasized by a black arrow. To take into account of this discontinuity, we use a Huber cost function for back surface smoothness.

becomes

$$\begin{aligned}\hat{\mathbf{t}}, \hat{\mathbf{l}} = \underset{\mathbf{t},\mathbf{l}}{\operatorname{argmin}} \sum_{c \in C} \|\mathbf{n}_{\mathbf{p},q}(t_c, l_c) - \mathbf{n}_{\mathbf{d},q}(t_c)\|_2^2 + \\ \lambda_1 \sum_{c \in C} \left\| l_c - \tilde{d}(c) \right\|_2^2 + \\ \lambda_2 \sum_{j,k \in N} \|t_j \mathbf{v}_{1,j} - t_k \mathbf{v}_{1,k}\|_2^2 + \\ \lambda_3 \sum_{c \in C} \left\| \frac{\partial}{\partial z} \mathbf{b}_j(t_c, l_c) \right\|_H , \end{aligned} \tag{5.9}$$

where $\mathbf{l}$ is a vector listing l_c for all pixels, j and k are pixel indices chosen from a set of all neighborhood N, $\|\cdot\|_H$ is the Huber penalty function [133][4], $\mathbf{v}_{1,c}$ is a unit camera ray vector corresponding to pixel c, and $\tilde{d}(c)$ is the measured ToF depth at pixel c. The first term ensures the normal consistency, the second term represents denoising of measured signal $\tilde{d}$, and the third and fourth terms regularize the denoising process by enforcing smoothness of front and back surfaces. If the front surface consists of multiple planes or is curved, continuity of back surface points could break as illustrated in Fig. 5.3. We therefore use a Huber cost function for the back surface smoothness because it allows occasional discontinuity (outliers) while retaining overall smoothness.

5.2.4 Solution Method

The optimization problem of Eq. (5.9) is unfortunately non-convex and difficult to directly solve. We therefore take an alternating minimization approach by splitting the original problem

[4] The Huber penalty function is an l_1/l_2 hybrid norm hence it can be used as smoothness with discontinuity, and defined as

$$\|\mathbf{x}\|_H = \sum_i h_\epsilon(x_i) \ \text{, where } \ h_\epsilon(x) = \begin{cases} |x| - \epsilon/2 & (|x| > \epsilon) \\ x^2/(2\epsilon) & \text{otherwise} \end{cases}$$

into two subproblems; one for estimating $\mathbf{t}$ with keeping $\mathbf{l}$ fixed ($\mathbf{t}$-subproblem), and the other for determining $\mathbf{l}$ with given $\mathbf{t}$ ($\mathbf{l}$-subproblem). This can be interpreted as alternating estimation of shape ($\mathbf{t}$-subproblem) and denoising of measurement ($\mathbf{l}$-subproblem). Our method iteratively updates $\mathbf{t}$ and $\mathbf{l}$ with initialization of $\mathbf{t} = \texttt{constant}$, which is chosen manually by setting the approximate distance to the object, and $\mathbf{l} = \tilde{\mathbf{d}}$, which is a vectorized $\tilde{d}(c)$ for all pixels. We now discuss these two subproblems in detail.

$\mathbf{t}$-subproblem By fixing $\mathbf{l}$, Eq. (5.9) can be reduced to an optimization problem of $\mathbf{t}$ as

$$\hat{\mathbf{t}} = \underset{\mathbf{t}}{\operatorname{argmin}} \sum_{c \in C} \left\| \mathbf{n}_{\mathbf{p},q}(t_c, \hat{l}_c) - \mathbf{n}_{\mathbf{d},q}(t_c) \right\|_2^2 + \lambda_2 \sum_{j,k \in N} \|t_j \mathbf{v}_{1,j} - t_k \mathbf{v}_{1,k}\|_2^2 , \tag{5.10}$$

where $\hat{l}_c$ represents the estimate of l_c obtained from the previous iteration. This problem can be interpreted as estimation of the front surface shape because t_c corresponds to the depth of front surface. Neglecting the back surface smoothness term is justified because of the fact that the length between $\mathbf{f}(t)$ and $\mathbf{b}(t)$ is almost fixed when $\hat{l}_c$ is unchanged. Due to the nested normalization terms as shown in Eq. (5.6), it is difficult to analytically derive its first and second-order derivatives of the objective function. To avoid high computational complexity of calculating second derivatives, we use the L-BFGS method [134] which only uses approximate Hessians rather than explicitly computing them. The $\mathbf{t}$-subproblem is again non-convex; therefore, there is a chance of being trapped by a local minima. As we will see in Sec. 5.3.1, in practice, it yields good estimates with appropriate initialization.

$\mathbf{l}$-subproblem By fixing $\mathbf{t}$ and neglecting the front surface normal consistency, we obtain an $\mathbf{l}$ optimization problem written as

$$\hat{\mathbf{l}} = \underset{\mathbf{l}}{\operatorname{argmin}} \sum_{c \in C} \left\| l_c - \tilde{d}(c) \right\|_2^2 + \lambda_3' \sum_{c \in C} \left\| \frac{\partial}{\partial z} \mathbf{b}_j(\hat{t}_c, l_c) \right\|_H , \tag{5.11}$$

where $\lambda_3' = \lambda_3 / \lambda_1$. It can be viewed as a problem of denoising $\mathbf{l}$ regularized by the back surface smoothness (the second term) with a fixed front surface $\hat{\mathbf{t}}$. Although this problem is not strictly convex, we observed that this problem is approximately convex[5]; hence the optimal solution can be obtained. We again use the L-BFGS method as a minimizer for $\mathbf{l}$-subproblem.

[5]Please refer to Appendix B.3 for the convexity of $\mathbf{l}$-subproblem.

The solution method for the baseline method described in Sec. 5.2.2 corresponds to the $\mathbf{t}$-subproblem. On the other hand, for the robust estimation method in Sec. 5.2.3, $\mathbf{t}$- and $\mathbf{l}$-subproblems are repeatedly solved in an alternating manner. In the implementation, we begin with the $\mathbf{t}$-subproblem and iterate until convergence. We stop the iteration when the gap between estimates and former estimates for both $\hat{\mathbf{t}}$ and $\hat{\mathbf{l}}$ are small enough.

5.3 Experiments

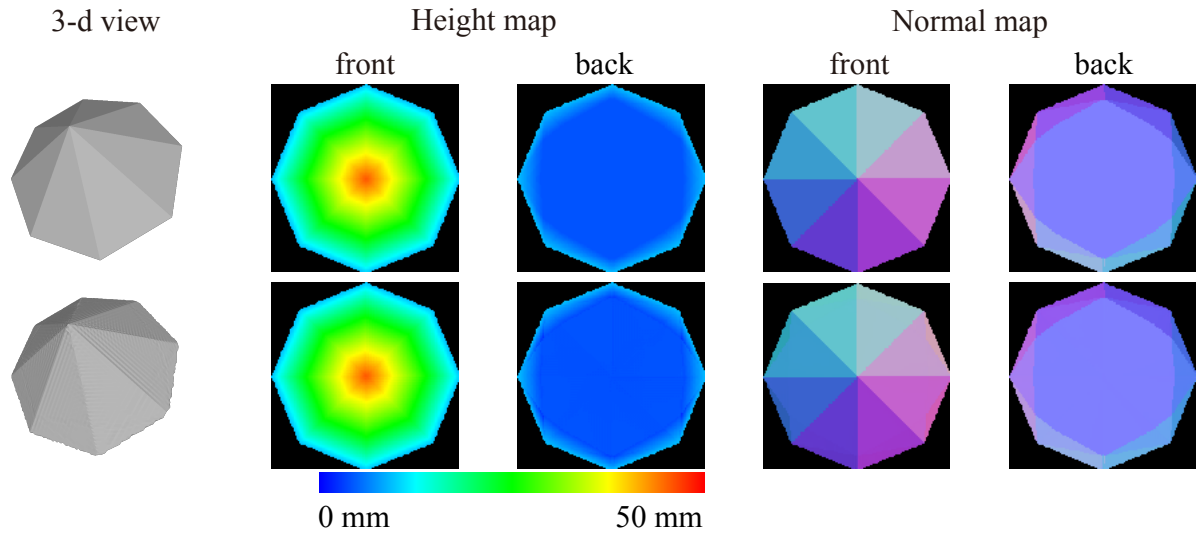
In this section, we first assess the accuracy, robustness, and convergence of the proposed method using simulation data, and apply the method to real-world transparent objects to evaluate its effectiveness.

5.3.1 Simulation Test

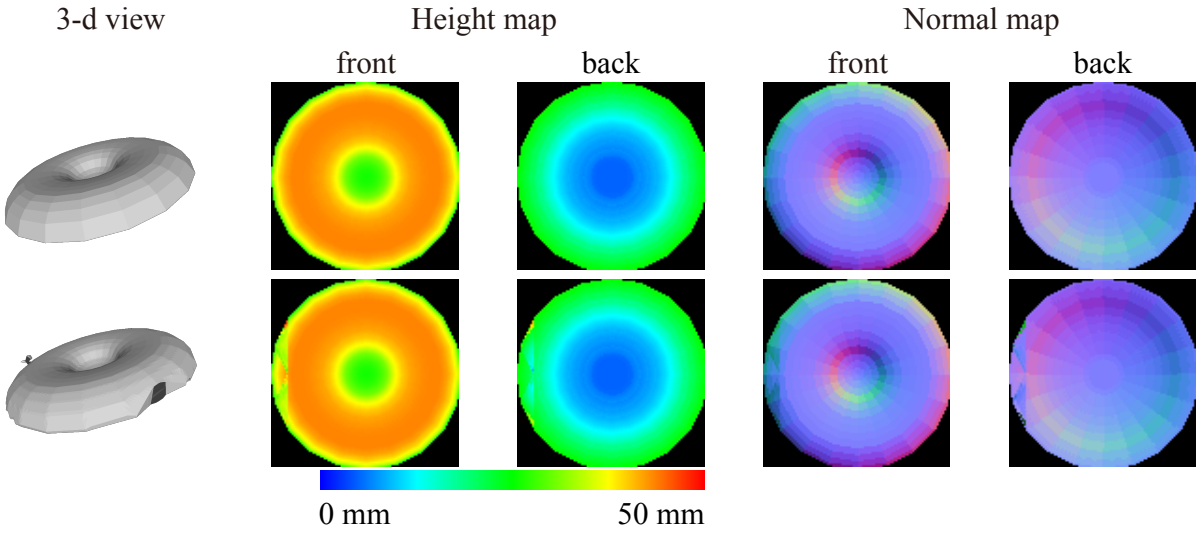
We generate simulation data of a scene with a transparent object that consist of the optical length and two reference points by ray tracing using Eq. (5.1) and Snell's law. To the data, we apply the proposed method to estimate front and back surface points and assess its accuracy by comparing with the ground-truth model. We use 0.005 for λ_2 and 20 for λ_3' for all target objects in the simulation tests.

Accuracy We first assess the effectiveness of the normal consistency objective (t-subproblem) using 48 types of transparent shapes generated by simulation. The data is noise-free in this experiment, and the solution is derived by solving the t-subproblem (baseline method with smoothness). The initial value of $\mathbf{t}$ is set to constant using an approximate depth to the object, which essentially corresponds to a planar surface located nearby the object. Figure 5.4(a) shows the result of a diamond shape scene. The result is close to the ground truth, whose root mean squared error (RMSE) is $0.524\,\mathrm{mm}$ (0.17 % of the optical length). As another example, Fig. 5.4(b) shows the result of a round torus-like object. The reconstruction accuracy is high in this case as well with RMSE $0.801\,\mathrm{mm}$ (0.26 % error), while there is visible artifacts at a few boundary regions. The average error of all 48 target objects is 0.45 %, and the result indicates the effectiveness of the normal consistency.

Effect of initialization We assess the influence of different initializations to the solution of t-subproblem because the t-subproblem is non-convex as described in Sec. 5.2.4. The transparent shape that we use for this test is placed at $200\,\mathrm{mm}$ from the camera, and its thickness is $50\,\mathrm{mm}$;



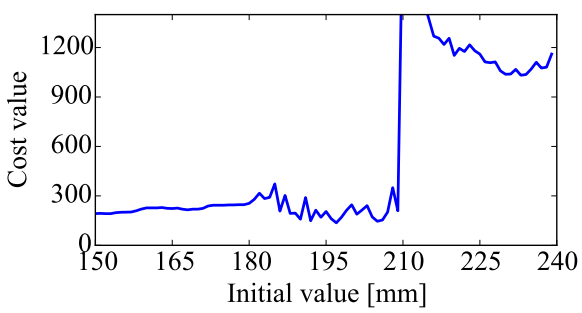
(a) Diamond object



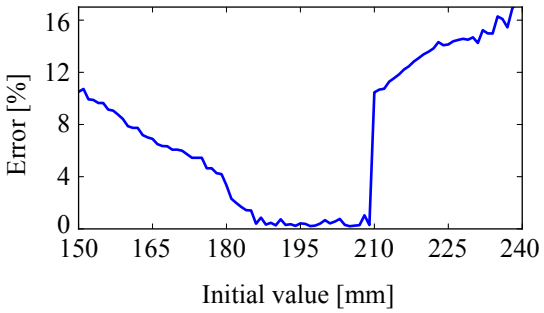
(b) Torus-like object

Figure 5.4: Simulation examples: Diamond and torus-like objects. The top row shows the ground truth of 3-d view, pseudo-colored height map, and normal map of both front and back surfaces. The bottom row shows the reconstruction result. The reconstruction error is 0.17% and 0.26%, respectively..

thus, the object spans in the range of $200\,\mathrm{mm}$ and $250\,\mathrm{mm}$. We vary the initialization constant for $\mathbf{t}$ from $150\,\mathrm{mm}$ to $250\,\mathrm{mm}$ and assess the convergence and accuracy by solving the t-subproblem. Figure 5.5(a) shows the cost value at convergence, and Fig. 5.5(b) shows the reconstruction error. The reconstruction error is less than 1% in a wide range of initial values between $186\,\mathrm{mm}$ and $209\,\mathrm{mm}$, which shows the tolerance of the method against inaccurate initial values.

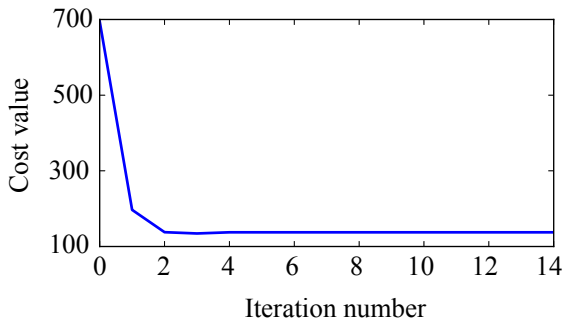


(a) Cost value

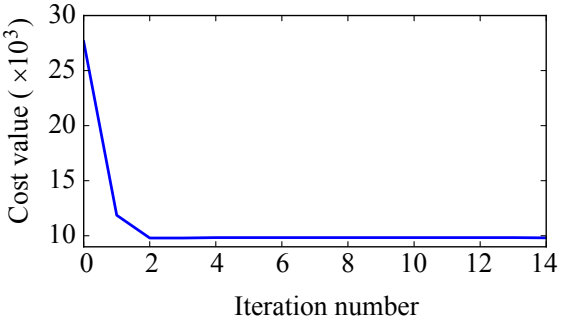


(b) Error of estimated shape

Figure 5.5: Assessment result about local minimum depending on the initial value. (a) Terminated cost value. (b) Error of estimated shape compared to the ground-truth. All initial values between $186\,\mathrm{mm}$ and $209\,\mathrm{mm}$ result in accurate shapes.



(a) Cost of t-subproblem



(b) Cost of l-subproblem

Figure 5.6: Convergence of our alternating minimization. (a) Cost of t-subproblem. (b) Cost of l-subproblem. They decrease over iterations and converge after 3 iterations in this example.

Convergence of alternating optimization We also assess the convergence of the alternating optimization described in 5.2.4. In this simulation, we add Gaussian noise with its standard deviation of 0.5% of optical length to all the pixels of the depth data. Figure 5.6 shows an example of the cost variations of two subproblems over iterations. The cost of both subproblems rapidly decreases at the beginning of iterations and remains stable after 3 iterations. With our test, most transparent shapes showed a similar convergence behavior and we consider that it is safe to say it reaches a local optimum.

Robustness against noise We further assess the effect of observation noise by adding noise to simulated ToF measurements. For comparison, we assess four approaches: (1) baseline method with smoothness (t-subproblem only), (2) t-subproblem with denoising, (3) alternating optimization (Sec. 5.2.4), and (4) alternating optimization with denoising. We use the Non-

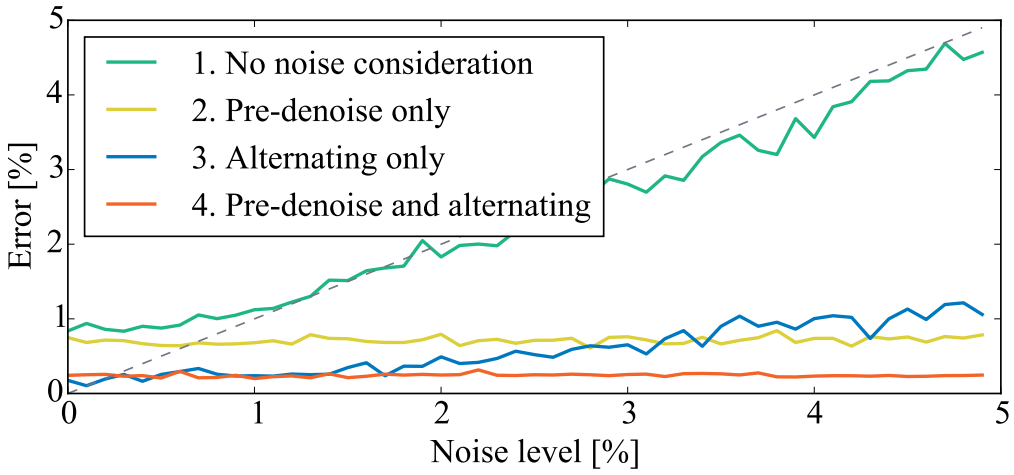


Figure 5.7: Reconstruction error of four approaches with respect to varying noise level. The reconstruction error generally becomes higher with a greater noise level. Errors are suppressed by proper noise handling. The combination of denoising and alternating optimization yields stable and the lowest error among these four approaches.

local Means Denoising [135] for denoising ToF measurements. Figure 5.7 summarizes the reconstruction errors of these four strategies with respect to the noise level. Without a proper noise handling (approach 1), results are significantly affected by observation noise hence the error increases together with the observation noise level. When denoising is applied prior to the optimization (approach 2), the error becomes stable regardless of the noise level. With the alternating optimization method (approach 3), the error is also suppressed while it shows degradation at higher noise levels. The highest accuracy is obtained by alternating optimization with denoising (approach 4). The result indicates that denoising is effective, that alternating optimization is also effective in suppressing noise, and that the combination of these two is effective. For the real-world experiments in the next subsection, we use the alternating optimization with denoising approach.

5.3.2 Real-world Experiment

For the real-world experiment, we use an off-the-shelf ToF camera (Kinect v2) and an LCD panel placed on a linear stage. Three distinct transparent objects are used for conducting the experiment. The parameters are set to $\lambda_2 = 0.05$ and $\lambda'_3 = 10$ for all target objects in this experiment.

Setup Figure 5.8 shows our experimental setup. We use Microsoft Kinect v2 for a ToF camera, whose lens is changed to Edmund Optics 35mm IR lens to narrower the field of view

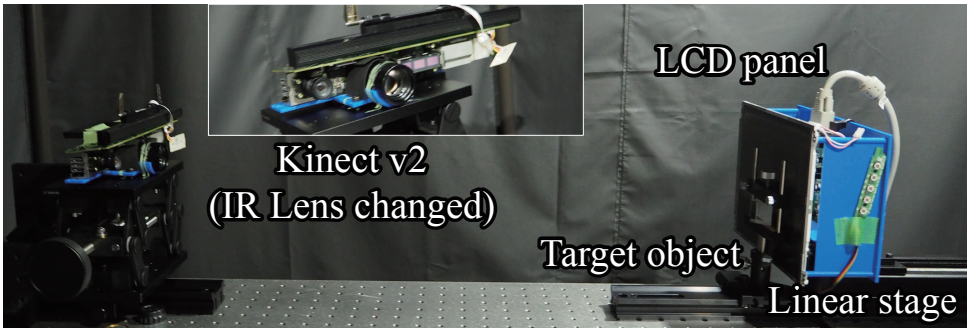


Figure 5.8: Experimental setup. IR lens of Kinect v2 is changed to obtain narrower field of view. Reference points are obtained using the LCD panel on a linear stage.

(FOV). To obtain reference points, we use an LCD panel, which is mounted on a motorized linear stage (OptoSigma SGSP26-150) that allows replication of positions in high precision. The LCD panel reflects the lights from Kinect with displayed patterns hence the back illumination of the display is turned off.

Calibration and measurements The LCD-camera system is calibrated at two LCD locations before measurement. The 3-d position of every LCD's pixels at two depth locations are measured in the form of IR and depth images of calibrated Kinect. The pixel location of the LCD panel is determined by Gray code pattern projection method [136]. For measurement, we place the target transparent object in between the camera and LCD panel. The target object is measured twice with distinct LCD panel locations. To obtain reference points in the measurement, we again use the Gray code pattern projection.

Result We conduct experiments using three transparent objects: cube, wedge prism, and schmidt prism, which do not produce multi-path interferences. Figure 5.9 shows target objects and the reconstruction results. We can see that estimated point clouds depicted as blue and yellow points well fit the ground-truth 3-d CAD model depicted as semi-transparent. To align the ground truth with the reconstruction, we use the Iterative Closest Point (ICP) [137] algorithm. Height and normal maps of both front and back surfaces are also shown. The quantitative reconstruction errors are summarized in Table 5.1. The mean error in the Euclidean distance and standard deviation of the error are small, and it shows the effectiveness of our method. The angle between two planar surfaces of a schmidt prism is 45 degrees, and calculated angle between planes from the recovered points, which is fit by PCA, is 45.10 degrees. Angle measures of the other objects are also efficiently small and it also shows the effectiveness.

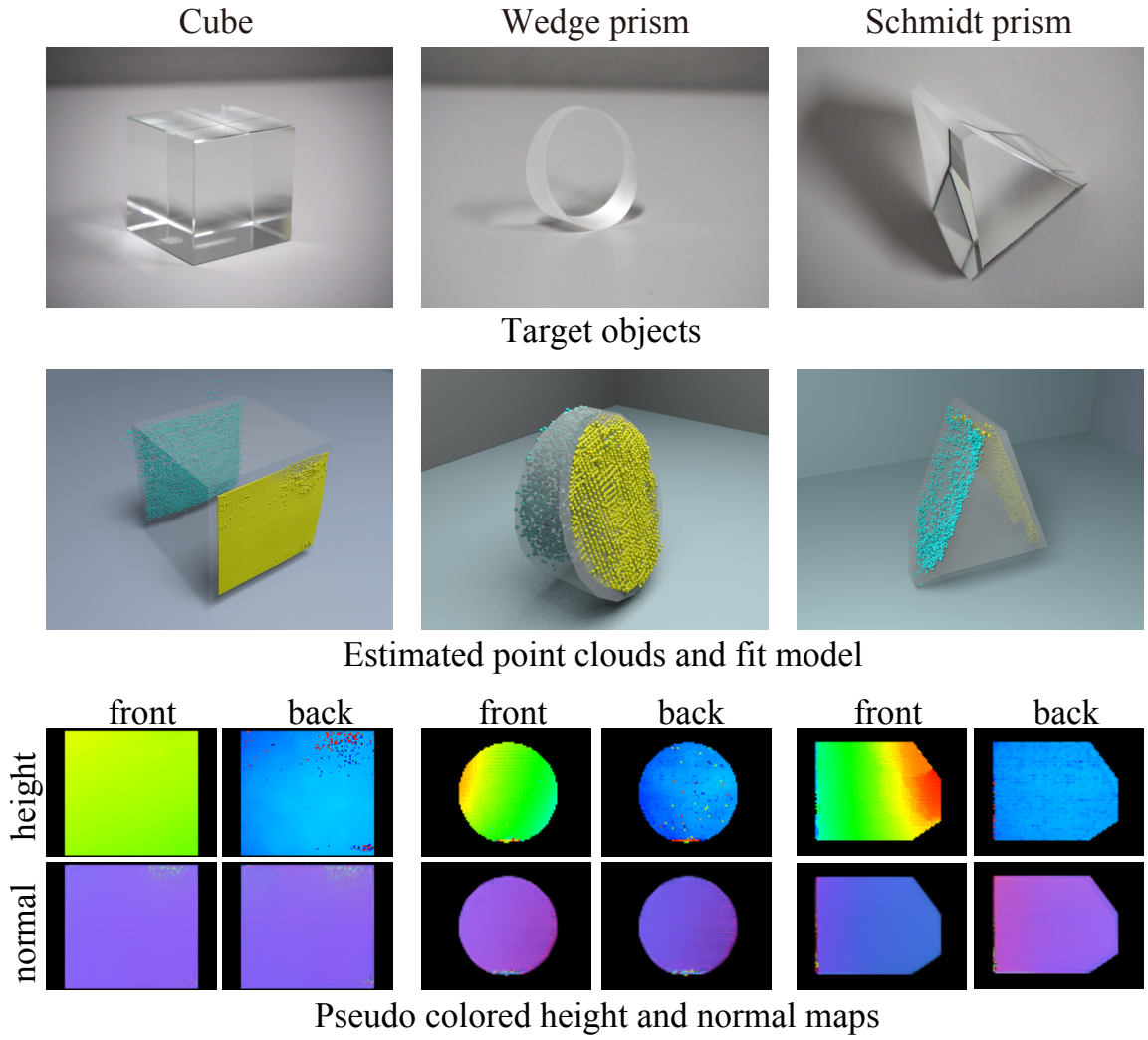


Figure 5.9: Experimental result of simple objects that consist of two planer surfaces. *Top:* Target objects. Front and back surfaces are single plane, hence no multi-path interferences are occurred. *Middle:* Estimated front points (yellow) and back points (cyan) with fit model by ICP (semi-transparent). *Bottom:* Estimated height map (upper) and normal map (lower) of front surface (left) and back surface (right).

Object	Mean	Std. dev.
Cube (parallel surfaces)	0.188 mm	0.458 mm
Wedge prism (18.8°)	0.226 mm	1.137 mm
Schmidt prism (45°)	0.381 mm	1.398 mm

Table 5.1: Quantitative evaluation of reconstruction errors. We evaluate the mean and standard deviation of the Euclidean distance between recovered points and the ground truth CAD model. The estimated points are registered to the model by ICP algorithm prior to the evaluation.

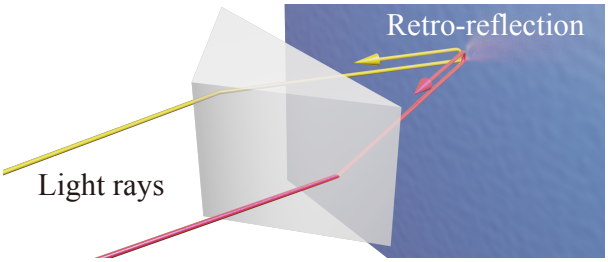
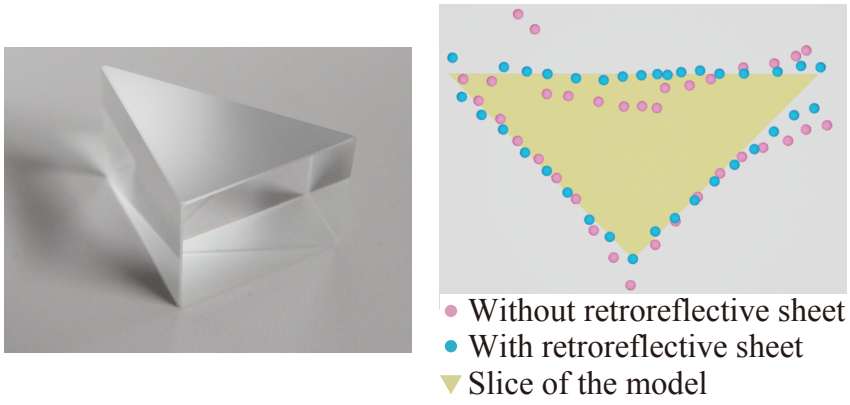


Figure 5.10: Multi-path avoidance using a retroreflective sheet. Using a retroreflective sheet, two rays (colored yellow and red) can be observed separately even when their reflection point is overlapped.

Multi-path avoidance If the target object is curved or has multiple planes, caustics appear on the background surface. In such a case, there are multiple light paths sharing the same reflection point on the background; therefore the observed depth by a ToF camera does not satisfy the model of Eq. (5.1). This effect, so called multi-path effect, is caused by the interference of multiple signals, where an ambiguously modulated wave is synthesized and the ToF camera measures it. To avoid this multi-path interference, we adopt and easy mitigation method using a retroreflective sheet placed on the background, which can be seamlessly combined to our measurement system. The retroreflective sheet reflects the incident light ray to the incident direction; hence, multi-path interference can be avoided even when the reference points are overlapped as shown in Fig. 5.10.

To verify the effectiveness of the multi-path avoidance, we use a right angle prism as shown in Fig. 5.11(a) and measure it with and without a retroreflective sheet. Figure 5.11(b) shows a slice of the estimated points and the ground truth model. Because the observation without a retroreflective sheet is affected by the multi-path interference effect, the result is distorted as depicted in the pink points. On the other hand, the result with the retroreflective sheet shows faithful recovery of the object as depicted as blue points because it mitigate the multi-path effect. It is also verified quantitatively by the reconstruction errors of the experiment summarized in Table 5.2.

Finally, we recover the shape of a convex lens, which has curved surfaces. We again use a retroreflective sheet for avoiding the multi-path effect. Figure 5.12 shows the estimated results. While it became somewhat noisy, the curved surface that exhibits multi-path refraction rays is recovered.



(a) Target object (b) Slice view

Figure 5.11: Reconstruction result of a right angle prism. (a) The target object. (b) a slice view of the estimated surface points and the ground truth. Blue and red points are the result with and without retroreflective sheet, respectively. The retroreflective sheet mitigates multi-path interference for more faithful recovery.

Object	Mean	Std. dev.
without retroreflection	0.745 mm	1.115 mm
with retroreflection	0.448 mm	0.828 mm

Table 5.2: Numerical evaluation of right angle prism. The result with retroreflective sheet is more accurate than that without it.

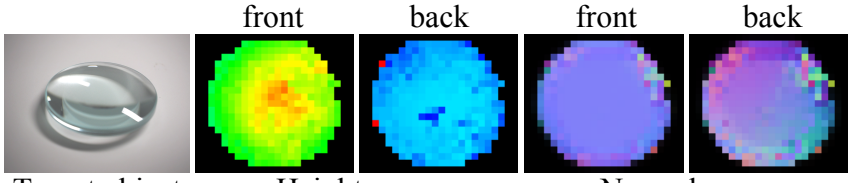


Figure 5.12: Reconstruction of a transparent curved object (convex lens)

5.4 Discussions

We developed a method for transparent shape recovery using the time-of-flight distortion. One of the issues of our method for a practical use is that it is currently limited to low-resolution, because it is bounded by the resolution of the ToF image sensor. Another issue is that our method breaks down when the light path refracts more than twice, *e.g.*, due to total reflections inside the transparent object, which could occur at the edge of the object. To avoid this problem, the region near such edges should be treated differently by developing a suitable technique.

The normal consistency objective appeared in (5.8) and (5.9) is defined as the Euclidean

distance of unit normal vectors. Theoretically, it should be better written by

$$\underset{\mathbf{t}}{\operatorname{argmin}} \sum_{c \in C} \left(\mathbf{n}_{\mathbf{p},q}(t_c)^T \mathbf{n}_{\mathbf{d},q}(t_c) - 1 \right)^2 ,$$

because of its directional nature. We have tested the above expression; however, the result did not change much while the computational cost increased significantly. Therefore, we decided to keep using the Euclidean distance of normals.

While we assume the neighbors of front surface points correspond to those viewed in the camera pixel coordinate, we cannot use the same assumption for the back surface because of refractions (depicted in Fig. 5.3), and a less restrictive smoothness term is defined for the back surface using a Huber cost function. We consider this design tends to make the front surface smoother than the back surface. This behavior can be balanced by adjusting weight parameters λs. Also, our solution is somehow heuristic approach for solving non-convex problem hence it is not guaranteed to reach a good estimate for any shapes.

In the future work, we are interested in exploring the direction of combining a computational illumination approach, such as [39], for not only mitigating the multi-path interference but also explicitly handling the multi-path effect for transparent shape recovery.

For reducing measurements, other optical cues are considerable. For example, amplitude decay by the Fresnel transmission and absorption convey the surface normal and thickness of the transparent object, and polarization change also conveys the surface normal of the object. Combining these optical cue and our proposed method, the number of measurements may be reduced or the accuracy can be improved. Considering these optical cue is also one of the interesting future direction of this research.

Chapter 6

Analysis of Temporally Spread Light and Time-of-Flight Distortion

In this chapter, we show an theory of Time-of-Flight distortion with varying frequencies. Using this theory, we propose a method for material classification indirectly using the difference of temporal light spread of material. We show that the time domain response of the material can be indirectly measured via the distortion of the Time-of-Flight camera and can be used for recognizing the material.

6.1 Preliminary

Material classification plays an important role for computer vision applications, such as semantic segmentation and object recognition. One of the major challenges in material classification is that different materials may yield very similar appearance. For example, artificial plastic fruits and real fruits confronting a camera produce visually similar RGB images that are difficult to distinguish. One of the possible strategies to distinguish similar appearance is to use the optical responses of the target object such as spatial, angular, and temporal spread of the incident light. Because different materials may have different optical responses due to their own subsurface scattering and diffuse reflection properties, it is expected that a more reliable material classification can be achieved using such optical cues on top of the RGB observations.

Recently, Heide *et al*. [36] have developed a method that recovers transient images from observations by a low-cost Time-of-Flight (ToF) camera, which is originally designed for depth measurement. There are other related studies that use ToF cameras for recovering ultra-fast light propagation, *e.g.*, impulse response, of the scene [39, 40, 38] with some hardware modifications and computation. Motivated by these previous approaches that exploit the temporal spread of light, we aim to classify materials using an indirect temporal cue from an off-the-shelf ToF camera without explicitly recovering impulse response.

(a) Mayonnaise bottle (b) Distorted depth (c) Classification (d) Corrected depth

Figure 6.1: Depth distortion of a ToF camera. (a) A mayonnaise bottle is measured by a Kinect. (b) Measured depth in a 3D view. There is a gap in depth between the mayonnaise and label regions. We use this depth distortion for material classification. (c) Material segmentation result. The material label is assigned for each pixel. (d) Application of material classification to depth correction. Depths are corrected based on the segmentation result and the distortion database. Depth gaps among materials are corrected and a faithful 3D shape is recovered.

We develop a material classification method based on a key observation that the measured depth of a translucent object becomes greater than the actual depth as shown in Fig. 6.1(b), where the depth gap between the mayonnaise and paper label regions is obvious. We show that this depth distortion is caused by the time delay due to subsurface scattering and varies along with both the modulation frequency of ToF camera and the distance between the target and the camera. Using the depth distortions as a *feature* of the material, we propose an exemplar-based material classification method.

The chief contributions of this paper are twofold. First, we demonstrate that the material classification is tractable by an off-the-shelf ToF camera, *e.g.*., Xbox One Kinect. Our method uses the distorted depth measurements as an indirect temporal cue for material classification without explicitly recovering impulse responses; therefore it does not require any modifications of hardware unlike [36, 39]. Second, we show how ToF measurements are distorted inside materials and along with depths. By moving the target object along the depth direction, rich information about the target can be obtained and it serves as important clue for realizing material classification.

6.2 Time-of-Flight Observation

To begin with, we briefly review the measurements that are obtained by a ToF camera. A correlation-based ToF camera illuminates a scene by the amplitude modulated wave $f_\omega(t)$ and measures its attenuated amplitude and phase delay. From the phase delay and the speed of light, the depth of the scene can be obtained.

In general, a scene can have the "multi-path" effect due to inter-reflections and subsurface scattering, which degrade the depth estimation accuracy. Image formation models regarding the multi-path effect have been well understood thanks to the previous works [36, 39, 59]; hence, we briefly explain one of the models that we are going to use in this paper. Following a phasor representation [59], amplitude and phase of the returned wave can be represented by a single complex value $\boldsymbol{c} \in \mathbb{C}$, called phasor, governed by the modulation frequency ω. The measured amplitude $\tilde{a}_\omega \in \mathbb{R}$ and depth $\tilde{d}_\omega \in \mathbb{R}$ of the ToF camera are obtained as

$$\begin{cases} \tilde{d}_\omega & = \frac{c}{4\pi\omega} \arg \boldsymbol{c}(\omega), \\ \tilde{a}_\omega & = |\boldsymbol{c}(\omega)|, \end{cases} \tag{6.1}$$

where the $\arg$ operator returns the angle of a complex phasor, and c is the speed of light.

When the illumination wave is a sine wave, *i.e.*, $f_\omega(t) = \sin(2\pi\omega t)$, the observed phasor can be represented as

$$\boldsymbol{c}(\omega) = \int_0^\infty r(t-\tau) e^{-2\pi i \omega t} dt, \tag{6.2}$$

where $\tau(>0)$ is the time of flight toward the surface of the object, $r(t)$ is the impulse response, or a point spread function (PSF), of the object along with the time t, and i is the imaginary unit. The impulse response is the summation of all possible paths $\rho \in \mathcal{P}$; therefore, $r(t)$ can be written as

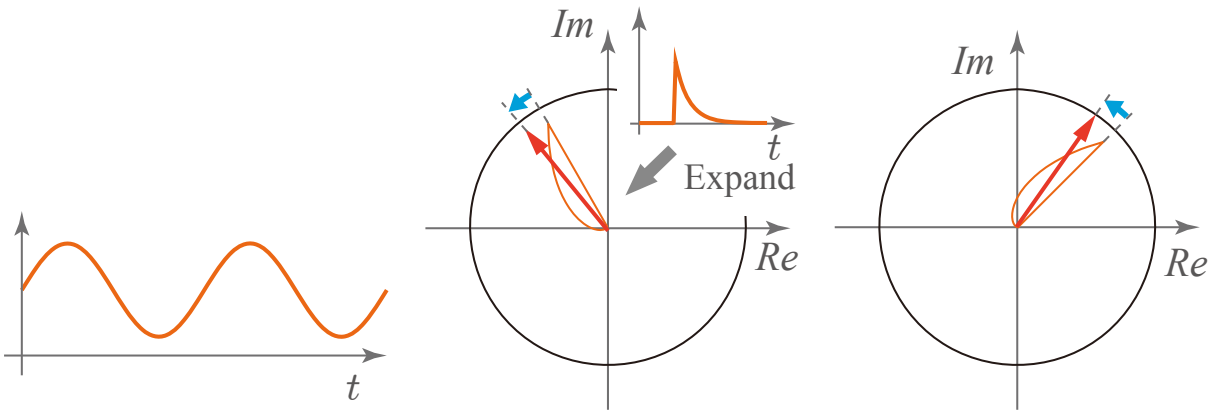
$$r(t) = \int_{\mathcal{P}} r_\rho \delta(|\rho| - t) d\rho, \tag{6.3}$$

where r_ρ is the contribution of the path ρ, $|\rho|$ is the time travelled along the path ρ, $\delta(t)$ is the Dirac delta function, and $t = 0$ is the time when the impulse light hits the surface of the object. Figure 6.2(b) illustrates a phasor representation of the multi-path ToF observation. The time domain PSF $r(t-\tau)$ is expanded onto the imaginary plane, and the phasor depicted by a red arrow is the integration of expanded PSF over the angle. Because the negative domain of $r(t)$ is zero, Eq. (6.2) expresses that ToF camera measures the Fourier coefficients of the impulse response at the frequency of the light modulation.

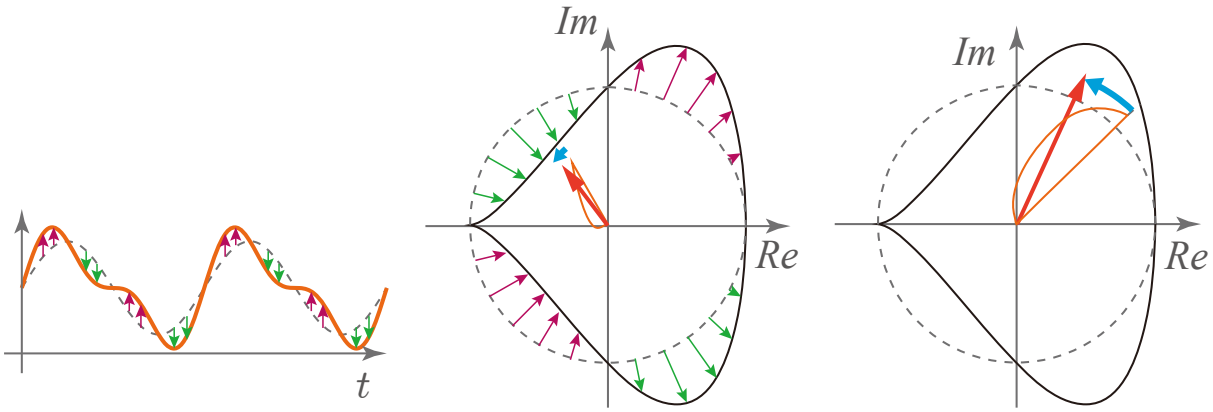
6.2.1 Frequency Dependent Depth Distortion

The principle of the ToF camera assumes that the impulse response forms Dirac delta function as $r(t) = \beta\delta(t)$, where β is the amplitude decay of modulated light. In this case, the measured depth $\tilde{d}_\omega$ becomes

$$\tilde{d}_\omega = \frac{c}{4\pi\omega} \underbrace{\arg \int_0^\infty \beta\delta(t-\tau) e^{-2\pi i \omega t} dt}_{=2\pi\omega\tau} = \frac{c\tau}{2} = d, \tag{6.4}$$



(a) Sinusoidal wave (b) Phasor representation (c) Phasor at different depth



(d) Distorted wave (e) Distorted unit ball (f) Large distortion of depth

Figure 6.2: Phasor representation of ToF observations. (a) Sinusoidal illumination, (b) Time domain PSF is expanded to the imaginary plane (orange). (c) When the object is placed at different depths, the observation gets rotated but phase distortion remains the same as (b). (d) Biased periodic illumination. This toy example adds 20% harmonics to the sinusoid for biasing. (e) The unit ball of the phasor representation is distorted due to the biased illumination. (f) The object is placed at the same depth as (c). The distortion of the phase becomes different than (e) and (c).

where $d = \frac{c\tau}{2}$ is the ground truth depth of the object. Equation (6.4) represents that the accurate depth can be obtained regardless of modulation frequency ω, if the impulse response of the scene is exactly the Dirac delta.

In reality, almost all materials except for the perfect mirror surface yield various shapes of impulse responses due to diffuse and subsurface scattering [58]. When the target object exhibits a temporally broad shape of the impulse response, band-pass characteristic in the frequency domain becomes unique to the object. Accordingly, ToF observation $\boldsymbol{c}(\omega)$ can take an arbitrary value, because $\boldsymbol{c}(\omega)$ is a Fourier coefficient of the impulse response $r(t)$ at the frequency ω. In such a case, $\arg \boldsymbol{c}(\omega)$ does not necessarily represent the correct phase $2\pi\omega\tau$, and as a result, the

measured depth $\tilde{d}_\omega$ is distorted, and the distortion varies with the modulation frequency ω. This frequency-dependent depth distortion is one of our key observations, and our method exploits this property for the goal of material classification.

The shift in the time domain corresponds to the shift of phase in the Fourier domain:

$$\begin{aligned}\mathcal{F}\left[r(t-\tau)\right] &= e^{-2\pi i\omega\tau}\mathcal{F}\left[r(t)\right] \\ &= e^{-2\pi i\omega\tau}\hat{r}(\omega),\end{aligned}$$

where $\mathcal{F}[\cdot]$ is the Fourier transform and $\hat{r}(\omega)$ is the Fourier transform of the function $r(t)$. Measured depth $\tilde{d}_\omega$ can then be represented as

$$\begin{aligned}\tilde{d}_\omega &= \frac{c}{4\pi\omega}\arg\left(e^{-2\pi i\omega\tau}\hat{r}\left(\omega\right)\right) \\ &= d + \frac{c}{4\pi\omega}\arg\hat{r}(\omega).\end{aligned} \tag{6.5}$$

The second term $\frac{c}{4\pi\omega}\arg\hat{r}(\omega)$ is the depth distortion at frequency ω. In Fig. 6.2(b), the depth distortion is illustrated as a blue arrow.

While a single observation of depth distortion can be the same among different materials by chance, multiple observations using varying modulation frequencies can be used for enriching the measurement. Such multiple observations can be obtained with negligible latency because the ToF measurement is much faster than the ordinary video frame intervals [38].

However in practice, it is not straightforward to measure distortions using many different frequencies by an off-the-shelf ToF camera. For example, Kinect has only three modulation frequencies, and the frequencies cannot be easily changed; hence, only three distortion measurements are practically available, which may be too few for developing a reliable material classification system. To increase the information about the material in an alternative and easy way, our method employs a strategy of changing the distance between the camera and object. Now, we discuss the *depth-dependency* of the depth distortion.

6.2.2 Depth Dependent Depth Distortion

When the target object is placed at a different depth $d + \Delta d$, $r(t-\tau)$ is shifted by $\Delta\tau = \frac{2\Delta d}{c}$ in the time domain. As a result, the measured depth $\tilde{d}'_\omega$ becomes

$$\begin{aligned}\tilde{d}'_\omega &= \frac{c}{4\pi\omega}\arg\left(e^{-2\pi i\omega(\tau+\Delta\tau)}\hat{r}\left(\omega\right)\right) \\ &= d + \Delta d + \frac{c}{4\pi\omega}\arg\hat{r}(\omega).\end{aligned} \tag{6.6}$$

The measured depth is just shifted by Δd, and the depth distortion $\frac{c}{4\pi\omega}\arg\hat{r}(\omega)$ remains the same as the one at the original position as in Eq. (6.5). Figure 6.2(c) illustrates the depth

distortion at a different depth in a phasor representation. The blue arrow, which represents the depth distortion, is the same as that of the original position as illustrated in Fig. 6.2(b).

So far, we have assumed that the illumination is a perfect sinusoidal wave. In practice, because a high-frequency sinusoidal wave is difficult to generate, today's ToF cameras emit non-sinusoidal periodic waves that contain high-order harmonics [138, 139]. When the illumination wave has harmonics components as shown in Fig. 6.2(d), the ToF observation exhibits depth-dependency as illustrated in Figs. 6.2(e) and 6.2(f). Let us suppose that the distorted sinusoidal wave is biased as $f_\omega(t) = b_\omega(2\pi\omega t)\sin(2\pi\omega t)$, where $b_\omega(2\pi\omega t)$ is a periodic bias of the illumination wave due to harmonics. The observed phasor is then written as

$$\boldsymbol{c}(\omega) = \int_0^\infty r(t-\tau)b_\omega(2\pi\omega t)e^{-2\pi i\omega t}dt. \tag{6.7}$$

The above indicates that the observation $\boldsymbol{c}(\omega)$ is the Fourier coefficient of $r(t-\tau)b_\omega(2\pi\omega t)$, where the impulse response $r(t)$ is distorted by the bias $b_\omega(2\pi\omega t)$. Obviously, the biased impulse response $r(t-\tau)b_\omega(2\pi\omega t)$ varies along with τ, *i.e.*, the observation varies along with the depth.

Usually, this depth-dependent variation is unwanted; therefore, previous works attempted to eliminate it. For example, Su *et al*. [123] remove the depth-dependent variation using a correlation matrix. In contrast, we use the depth-dependent distortion as an important cue for material classification as it contains rich information about the target's response.

6.3 Material Classification

Our method uses either or both of the frequency- and depth-dependent depth distortions of ToF observations for the purpose of material classification. For describing how to use the depth distortions for material classification, we begin with the case where the actual depth is known and later describe a more general case where such an assumption is eliminated.

When the target object is placed at a known depth location, the depth distortion with respect to the actual depth is directly measurable. Let us suppose that the target object is measured by $n(\geq 1)$ modulation frequencies and $m(\geq 1)$ positions. The absolute depth distortion a_{ω_i,d_j} can be obtained by

$$a_{\omega_i,d_j} = d_j - \tilde{d}_{\omega_i,j}, \tag{6.8}$$

where $\tilde{d}_{\omega_i,j}$ is the measured depth at the i-th modulation frequency ω_i ($i \in \{1,\cdots,n\}$) and the j-th position ($j \in \{1,\cdots,m\}$), and d_j is the actual depth at the j-th position. By aligning

these distortions, a mn-length vector $\mathbf{v}$ can be formed as a *feature* vector of the object as

$$\mathbf{v} = \begin{bmatrix} a_{\omega_1,d_1} & \cdots & a_{\omega_n,d_m} \end{bmatrix}^T. \tag{6.9}$$

Because the actual depth of the target object is not generally accessible, we develop a feature that does not require the knowledge of the actual depth. Although we cannot directly obtain the depth distortion in this case, the *relative* depth distortions among multiple frequencies and/or multiple depths can be alternatively used. When multiple modulation frequencies are available, *i.e.*, $n \geq 2$ case, the relative depth distortion a'_{ω_i,d_j} can be computed by regarding the measurement of one of the modulation frequencies, say the n-th modulation frequency, as the reference measurement. The relative depth distortions can be obtained by taking differences from the reference measurement as

$$a'_{\omega_i,d_j} = a_{\omega_i,d_j} - a_{\omega_n,d_j} = \tilde{d}_{\omega_n,j} - \tilde{d}_{\omega_i,j}, \tag{6.10}$$

where i ranges from 1 to $n-1$. We can then setup an $m(n-1)$-length vector $\mathbf{v}$ by aligning the relative depth distortions, and it can be used as a feature vector for material classification. Although the reference measurement $\tilde{d}_{\omega_n,j}$ may be distorted depending on the material, the feature vector $\mathbf{v}$ encapsulating the relative distortions conveys discriminative cues for classifying materials.

In a similar manner, for the case where a single modulation frequency and multiple depth locations is available, *i.e.*, $n = 1$ and $m \geq 2$, the relative depth distortions among depth locations a''_{ω_1,d_j} can be obtained by regarding the measurement of the m-th depth position as the reference measurement as

$$a''_{\omega_1,d_j} = a_{\omega_1,d_j} - a_{\omega_1,d_m} = \tilde{d}_{\omega_1,m} - \tilde{d}_{\omega_1,j} + \Delta d_j, \tag{6.11}$$

where Δd_j is the amount of movement from the base position, which should be measured.

6.3.1 Classifier

We assume that we have a database of materials that consists of the feature vectors measured using predefined modulation frequencies and depth locations in a certain range beforehand. For classification, the target object is measured with the full or partial set of the predefined modulation frequencies and depth locations. Once we obtain the feature vector of the target object as a query, we use the material database as exemplar to look up the closest material.

While any arbitrary classifiers can also be alternatively used, it is desired for classifiers to have the following two properties. First, since the feature vectors tends to be high-dimensional

while the number of materials in the database may be small, it is preferred the classifier uses a model with a small number of parameters, or non-parametric like our choice. Second, a capability of handling missing elements in the feature vector is practically important, because the measurement is sometimes missing due to specular reflection on the object surface, or becomes saturated with near-distance reflectance.

For these reasons, we adopt a simple nearest neighbor classifier, which assesses the Euclidean distance (ℓ_2 norm). To deal with the missing or uninformative saturated observations, we remove such elements in the feature vector when evaluating the distance. The distance ξ_p between the feature vector $\mathbf{v}$ of the target object and the training vector $\mathbf{v}^p$ of the object p in the dataset can be computed as

$$\xi_p = \frac{1}{N} \sum_{k=0}^{nm} \begin{cases} 0 & v_k = \text{N/A} \\ (v_k - v_k^p)^2 & \text{otherwise,} \end{cases} \tag{6.12}$$

where N is the number of valid elements, and v_k and v_k^p are k-th element of vectors $\mathbf{v}$ and $\mathbf{v}^p$, respectively. Using this distance, we can classify the object by searching the nearest class $\hat{p}$ as

$$\hat{p} = \operatorname*{argmin}_p \xi_p.$$

Throughout the evaluation in this paper, we use this simple nearest neighbor classifier to assess the effectiveness of the depth distortion features for material classification.

6.4 Experiments

We evaluate the proposed method by a ToF camera and a linear translation stage system as shown in Fig. 6.3. We use Microsoft Kinect v2 for a ToF camera, which has three modulation frequencies ($n = 3$), and a OptoSigma's translation stage (SGSP46-800). As the official Kinect API does not support an access to depth measurements of each frequency, we have slightly altered an open-source software `libfreenect2`[1] to obtain such data.

First, we measure the depth distortion data for many materials and examine their differences across materials. The target object is placed on a linear translation stage changing the depth from 600 mm to 1250 mm ($m = 2600$), and is measured several times with changing the orientation of the object. The ground truth depth is obtained from the position of the translation stage, which is only used this test. Figure 6.4 shows the depth distortion of three materials; white acrylic board, polystyrene board, and opal diffusion glass. They are visually similar

[1] `https://github.com/OpenKinect/libfreenect2`

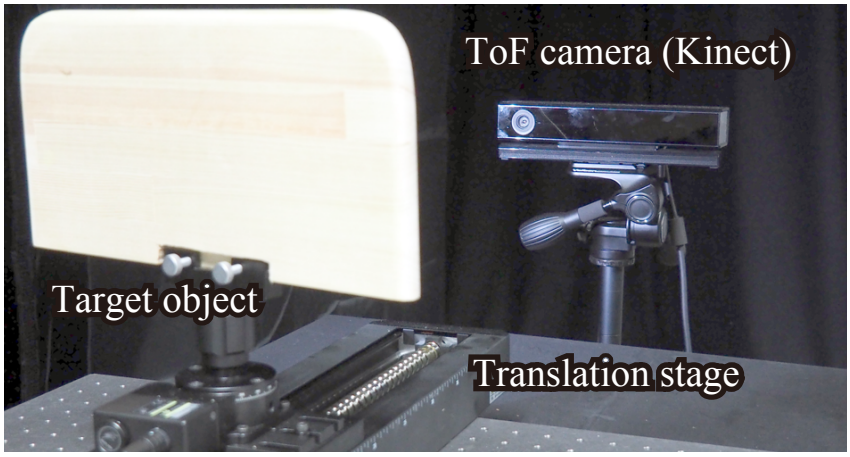
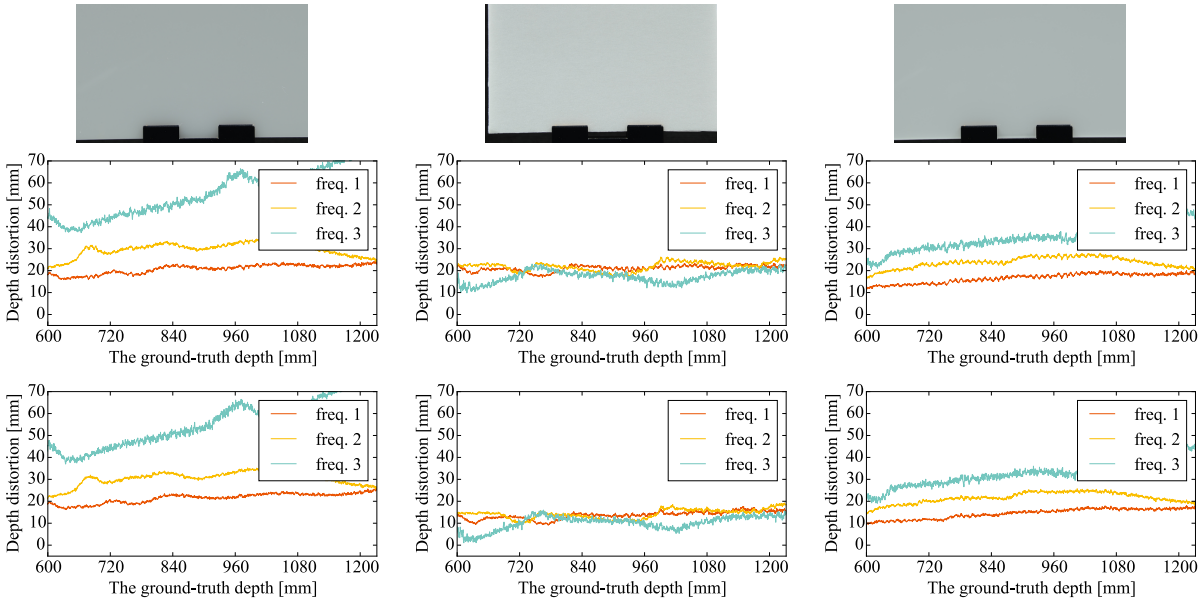


Figure 6.3: Experimental setup. We use Kinect as a ToF camera, and the target object is placed on a linear translation stage.



(a) White acrylic board (b) White polystyrene board (c) White diffusion glass

Figure 6.4: Measured depth distortions using Kinect for three objects. The ground truth depth is obtained via a linear translation stage. The top row shows photographs of the target objects. Measurements of the second and third rows are different in terms of surface orientation. Depth distortion of each frequency varies along with the actual depth and material. Depth distortion is similar for the same material regardless of the surface orientation, but largely different in different materials. This frequency- and depth-dependent depth distortion is our key observation for material classification.

object (white, planer, and no texture) hence appearance based methods have difficulty to distinguish these objects. On the other hand, depth distortions of ToF observations show significant difference across materials and retain consistency over measurement sessions.

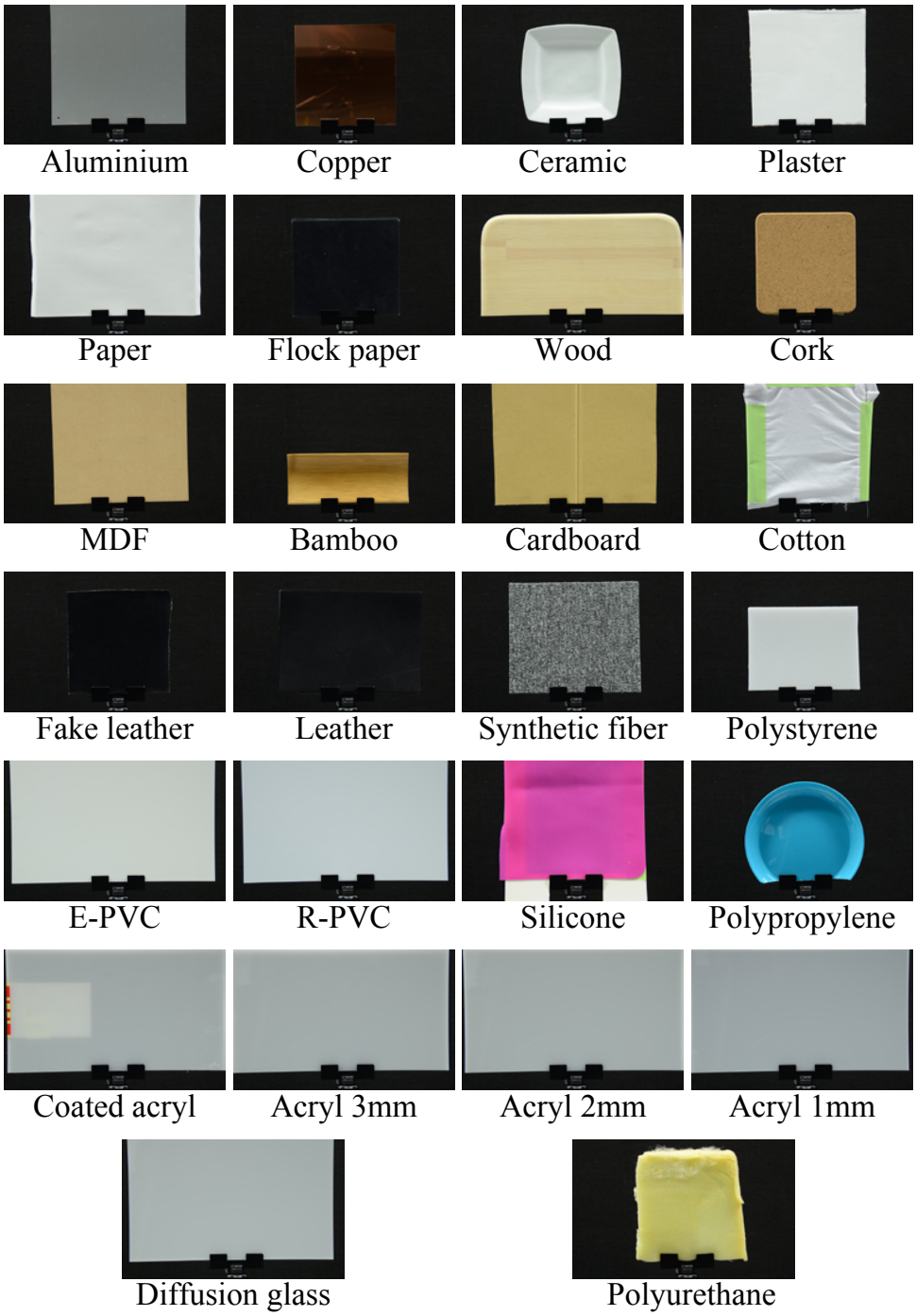
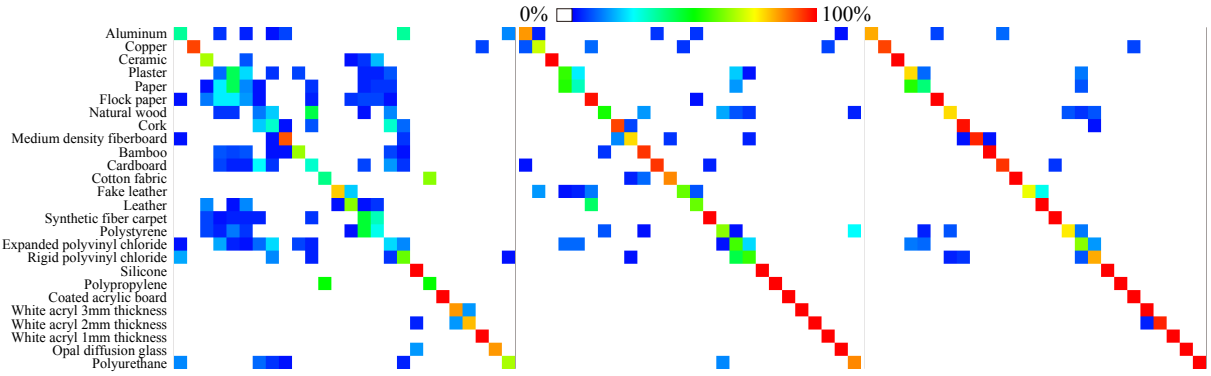


Figure 6.5: All materials of our database. All images are captured by the same camera parameters *e.g.*ISO, f-number, shutter speed, and focal length.

Using this depth distortion data, we assess the accuracy of material classification by the nearest neighbor classifier. The dataset consists of 26 materials including metal, wood, plastic, fabric, and so on as shown in Fig. 6.5, with 13 orientations for each material to enable the classifier to deal with target objects with arbitrary surface orientations. We evaluate the classification



(a) Frequency-dependent distortion only

(b) Depth-dependent distortion only

(c) Both of them

Figure 6.6: Confusion matrix. Red indicates the higher value and it appears on the diagonal. (a) Classified result only using frequency-dependent distortion. Accuracy is 57.4%. (b) Classified result only using depth-dependent distortion. Accuracy is 81.6%. (c) Classified result using both of them. Overall accuracy is 90.5%.

accuracy using three different features: Frequency-dependent distortion, depth-dependent distortion, and both of them. Using the feature with only frequency-dependent distortion ($n = 3$ and $m = 1$), the accuracy is 57.4%. This low accuracy is due to the limited availability of the number of frequency channels. Using only depth-dependent distortions ($n = 1$, $m = 2600$, and using Eq. (6.11)), the accuracy is improved to 81.6%. Finally, with both of frequency- and depth-dependent distortions ($n = 3$ and $m = 2600$), the accuracy is further improved to 90.5%. The confusion matrices are shown in Fig. 6.6. While many materials are correctly classified, some materials are miss-classified. For example, plaster and paper, or expanded and rigid polyvinyl chlorides have similar impulse responses due to similar scattering properties; therefore they are sometimes miss-classified.

Feature variations w.r.t surface orientation When the surface orientation of the target object varies, the time-domain impulse response may also vary. To confirm the effect of surface orientations, we measure a wooden board by changing the orientation and assess the variation of feature vectors with respect to varying orientations. Figure 6.7 shows the variation of the nearest distance from the wood class in the feature space along with the surface orientation of the target object. The red line indicates the upper-bound distance from the wood class, under which the query feature vector is correctly classified as "wood." In other words, once the distance from the wood class to the query feature goes beyond this upper-bound distance, it will be misclassified. The feature is stable under around 70 degrees, which indicate that the depth

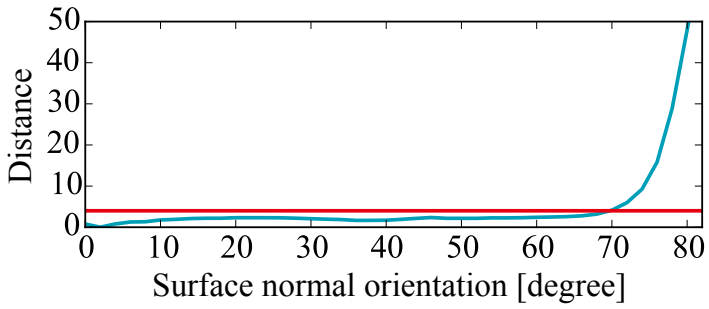
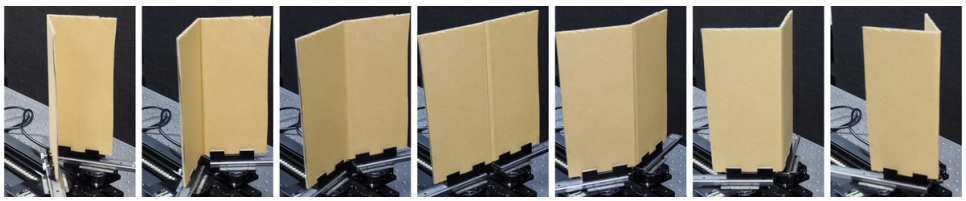
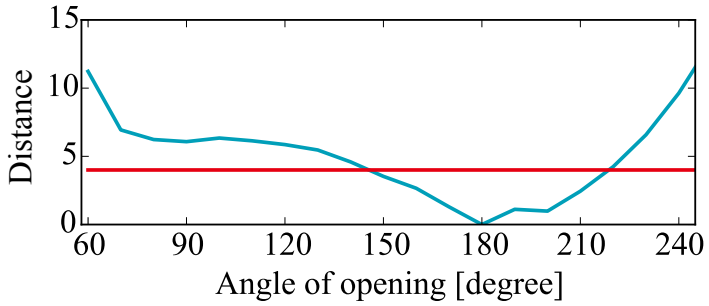


Figure 6.7: Feature vector variation over surface orientation. We change the orientation of the target object, and plot the distance of features along with the orientation. The feature is stable under around 70 degrees, and shows large deviation at steep-slant orientation. Red line indicates the upper-bound distance for the correct classification.



(a) Scenes of the folded cardboard

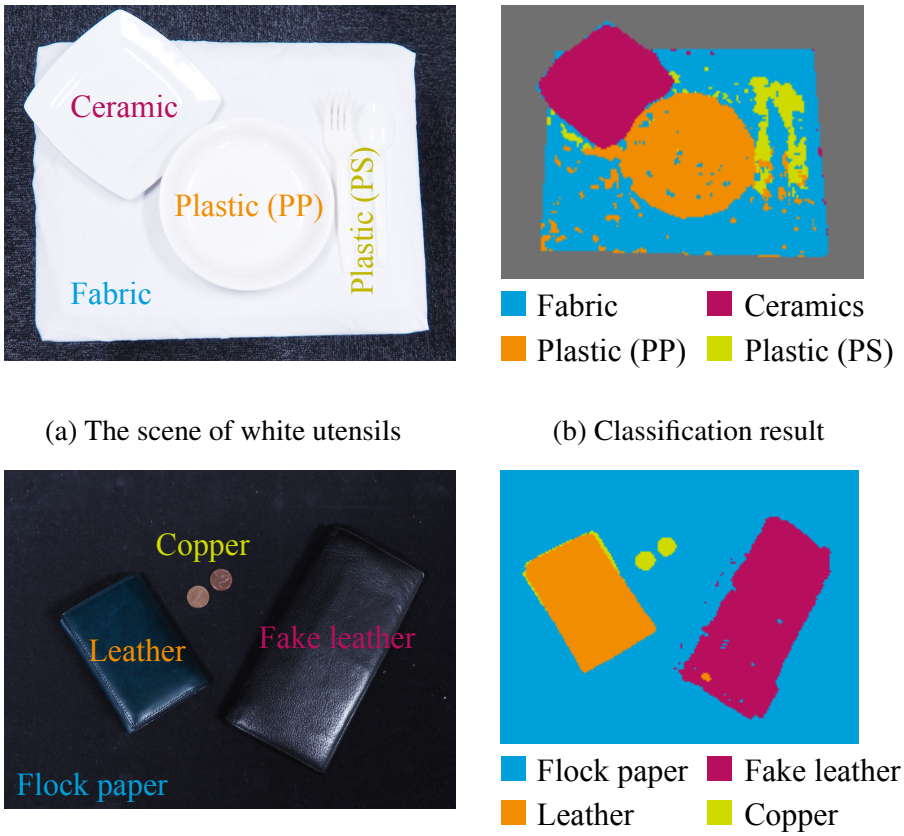


(b) Variation of feature distance between flat and folded cardboards

Figure 6.8: Shape dependency of the feature vector. We measure a cardboard with folding from 60 to 240 degrees. By folding cardboard less than 180 degrees, the scene exhibits strong inter-reflections.

distortion feature is reliable for the confronting surface in practice but may break down for a steep-slanted surface, *e.g.*, near the edges of a round-shape object.

Feature variations w.r.t. shape When the shape of the target object varies, the time domain impulse response may also vary, especially for a concave shape where significant inter-reflections occur. To confirm the effect of the shape of the object, we set up a scene of folded cardboard that can change the opening angle. We measure the folding edge area of the cardboard with changing the opening angle from the small angle (closed) via 180 degrees (flat) to



(a) The scene of white utensils

(b) Classification result

(c) The scene of wallets

(d) Classification result

Figure 6.9: Material segmentation results. (a) All utensils are white hence it is difficult to classify only with an RGB image. (b) The result of our material classification. Although there are some estimation error because of the pixel-wise process and only one depth variation, the scene is much interpretable than the RGB image. (c) Wallets made of genuine and fake leather and copper coins are placed in the scene. (d) Material segmentation result.

large angle (protruded) as shown in Fig. 6.8(a). The distances of feature vectors between the folded and flat cardboards are plotted in the blue line in Fig. 6.8(b). The red line represents the upper-bound of the flat cardboard class, under which the target is regarded a flat cardboard, and a moderate robustness against the shape variation is shown.

Material segmentation Our method can be applied in a pixel-wise manner to achieve material-based segmentation. Figure 6.9 shows a couple of example of material segmentation. For the scene shown in Fig. 6.9(a), all objects in the scene are white and the material classes are not obvious in the RGB image. With our method, the material is classified for each pixel as shown in Fig. 6.9(b). For this application, we use only frequency-dependent variations without the depth-dependent ones, *i.e.*, $m = 1$, because the alignment of the pixels may become hard when

the geometric relationship between the camera and scene varies. As a result, the result appears to be a little bit noisy, but it still shows faithful classification performance. For this experiment, we used a reduced database containing only four materials as the dimensionality of the feature vector is limited. Figure 6.9(c) shows another scene where wallets made of genuine and fake leather are placed, and they are correctly classified as shown in Fig. 6.9(d).

Depth correction Once materials are classified, the distorted depths can be corrected for recovering an accurate 3D shape using the material database that contains the samples of distortions for all materials. An example of the depth correction is shown in Fig. 6.1(a). Because mayonnaise has significant subsurface scattering, the measured depth of mayonnaise region is strongly distorted than that of the label as shown in Fig. 6.1(b). Figure 6.1(c) shows our result of material segmentation. Again, we do not change the depth of the target; therefore, only frequency-dependent variation is used ($m = 1$) with a limited database. Although some artifacts are observable because of the limited amount of measurement and steep surface orientations, mayonnaise and the label regions are largely well separated. Using the segmentation result and depth distortion database, a faithful 3D shape of the mayonnaise bottle is recovered as shown in Fig. 6.1(d). Compared to the original shape, the depth discontinuity between mayonnaise and the label regions is significantly reduced.

Real-time classification system We develop a near real-time material classification system, which can recognize the target material category by a hand-held ToF camera. Using the partial matching described in Eq. (6.12), our method outputs the result in near real-time even when observations at only a small number of depth locations m is available. By increasing the variation of depths by moving a target object or the camera (increasing m), the classification accuracy is gradually improved because richer information can be fed to the classifier.

Thickness classification Depth distortion is also useful for thickness estimation of the optically thin material. For example, white acrylic boards are optically thin so that the impulse response varies along with its physical thickness. The thickness of the white acrylic board can therefore be classified as shown at the near bottom part of Fig. 6.6. Currently, our method is limited to classification of different thicknesses, but we are interested in turning the problem into a regression problem for estimating the thickness.

6.5 Discussions

We have developed a material classification using an off-the-shelf ToF camera. We show that the measured depth using a ToF camera is distorted according to the time domain impulse response of materials, and the distortion varies along with the modulation frequency and the distance between the object and the camera. We use the ToF depth distortion as a cue for material classification, and developed a classification method.

Our method is based on a difference of time domain impulse response among materials, hence we assume the impulse response is the same for the same material. However, it may not be always true because complex shapes may cause varying impulse response. We have assessed the variation of the designed feature over varying shape and surface orientation and have shown the robustness of the developed feature up to a limitation on the variations. Related to this problem, optically thin object's impulse response also varies along with the object's thickness. On one hand, this allows us to classify thickness of the target object, but on the other hand, it indicates that a database with varying thicknesses is needed for correctly classifying materials of an object that may have arbitrary thickness. This is one of the current limitations of our method.

Another limitation is that the depth distortion measures, especially the depth-dependent distortion, is camera-dependent because the bias of illumination wave may be different across different devices. The development of the inter-device feature or transferring the database for a different camera is an important future work.

The amplitude of ToF observation also varies over different frequencies and depths, hence it can be also used for analyzing the scene. We did not use this cue in this paper, but we are interested in investigating this respect for further improving the classification accuracy.

Chapter 7

Conclusion

We established model-based approaches for measuring translucent and transparent objects using spatially and temporally modulated light. Amplitude-modulated illumination techniques convey rich information of the PSFs, and appropriate image processing methods can recover informative images from degraded observations. We developed high-frequency illumination and image processing methods based on the model of the spatial spread of light. We also developed a method that processes ToF distortions based on the model of the temporal spread of light.

Our aim was to avoid redundant measurement of PSFs. Instead we used the frequency responses of PSFs under amplitude-modulated illumination at different frequencies. Specifically, we

- revealed the applicability of high-frequency illumination for separating spatially spread light in the viewpoint of the light path design;
- analyzed the relationship between the spatial PSFs and the frequencies of projection patterns;
- analyzed the light path of ToF and the change in speed of light inside the medium; and,
- proposed a model of frequency- and depth-dependent ToF distortion for material classification.

Based on calculations using the relationship between the PSF and the frequency responses, we recovered invisible information from the scene.

In Chapter 3, we proposed a descattering method of transmissive observation, where the transmissive and scattered light are separated by parallel high-frequency illumination. We explained the principles of the high-frequency illumination method and the importance of avoiding overlapping rays. We proposed an optical design for transmissive observation to preserve

the correspondence between the projector and the camera. We developed a projector–camera system using telecentric lenses for our proposed method, and confirmed the effectiveness of the sharpening process in several real-world experiments, numerical evaluations, and living-body applications.

In Chapter 4 we described a method for recovering inner slices of translucent objects based on multi-frequency pattern projection. The proposed method is built upon the observation that the PSF inside translucent objects varies according to the depth within the object. Thus, we showed that inner layer can be recovered by estimating PSFs using different projection patterns. We also developed a method for automatically selecting informative slices via a sparse representation, by determining sparse coefficients that correspond to radiance slices. We further extended the method for dealing with inhomogeneous translucent objects based on a combination of pixel-wise appearance recovery and gradient-based image repairing. We verified the effectiveness of the proposed method using simulation experiments and real-world translucent objects.

In Chapter 5, we developed a method for transparent shape recovery using ToF distortions. Using the refractive index, we determined a unique refractive path using surface normal consistency. In addition, we showed a simple multi-path mitigation for this method using a retro-reflective sheet. We evaluated the effectiveness of this method using simulation and real-world experiments.

In Chapter 6 we described a material classification using an off-the-shelf ToF camera. We showed that the measured depth using a ToF camera is distorted according to the time-domain impulse response of the material, and that the ToF distortion varies with the modulation frequency and with the distance between the object and the camera. We used the ToF depth distortion as a parameter of the material, and classified numerous materials with high accuracy.

7.1 Limitations and Future Work

Our techniques are designed for each specific purpose, limiting the eligible target scenes. This is because our optical designs and computational algorithms are based on the specialized light transport models for the target applications. Different measurement and computation techniques based on the model of target's light transport must be developed, to visualize different information of the target object.

The exact cause of ToF distortions remains unclear. In the limited condition, ToF distortions are due to the material ' s refractive index, as described in Chapter 5, and translucency, as

described in Chapter 6. ToF distortions may also be affected by the shape, motion, density, temperature, and other object properties. Clarifying the effects of these properties and establishing a framework for all ToF distortions is an exciting topic for future research.

Proposed techniques include dealing with the behavior of light transport separately. Concurrently analyzing the spatiotemporal spread of light will provide more detailed and accurate analysis of the scene. Moreover, angular spread, spectral spread, and coherence spread, including polarization and interference, should also be included in future analysis. Developing a framework to incorporate all optical cues is another interesting topic for future research.

Appendix A

Hardware

To encourage the reproductivity, we summarize our hardware list in this appendix.

A.1 Hardware for Parallel High-frequency Illumination

List of parts that we used for the setup described in Chapter 3 is summarized in Table A.1. To align the optical axes, we mount a mirror onto the stacked gonio stages.

#	Part description	Model name	Manufacturer
1	Projector	LightCommander	Texas Instruments
2	Camera	Grasshopper2 GS2-FW-14S5M-C	Point Grey
3	Telecentric lens	#55-348	Edmund Optics
4	Telecentric lens	#62-922	Edmund Optics
5	Telecentric lens mounter	#63-440	Edmund Optics
6	Telecentric lens mounter	#56-027	Edmund Optics
7	Mirror	#48-453	Edmund Optics
8	Breadboard	#54-641	Edmund Optics
9	NIR polarizer	#48-894	Edmund Optics
10	Polarizer holder	#55-455	Edmund Optics
11	Mirror mount	KMH-80	OptoSigma
12	Gonio stage	GOHT-60A85B	OptoSigma
13	Gonio stage	GOHT-60A60B	OptoSigma
14	Right angle blacket	LBR-6085	OptoSigma
15	Optical bench	OBT-500SH-SCALE	OptoSigma
16	Optical bench career	CAA-60LS	OptoSigma
17	Optical bench career	CAA-25LS	OptoSigma
18	Base plate	BSP-70120 and BSP-2460	OptoSigma
19	Rod stand	BRS-12-60	OptoSigma
20	Rod	RO-12-120	OptoSigma
21	Cross clamp	CCHN-12-12	OptoSigma
22	Target platform	CHA-60	OptoSigma

Table A.1: List of parts for Parallel High-frequency Illumination system.

A.2 Hardware for Multi-frequency Illumination

List of parts that we used for the setup described in Chapter 4 is summarized in Table A.2. The camera is fixed on the optical bench, the projector is mounted onto the XYZ translation stage (consisted by XY stage and jack), and the beam splitter is mounted onto the rotation stage. The optical axis is adjusted by the projector's position and mirror angle.

#	Part description	Model name	Manufacturer
1	Projector	LightCommander	Texas Instruments
2	Camera	Aspen CG-6	Apogee
3	F-mount lens	Ai Micro-Nikkor 105mm f/2.8S	Nikon
4	Camera jig	Cut duralumin	–
5	Projector jig	Cut duralumin	–
6	Beam splitter for NIR	#48-905	Edmund Optics
7	Beam splitter for Vis	#48-904	Edmund Optics
8	XY transpation stage	#55-022	Edmund Optics
9	Lab. jack	#54-687	Edmund Optics
10	Right angle blacket	#55-380 and #55-381	Edmund Optics
11	Beam splitter holder	KMH-80	OptoSigma
12	Rotation stage	KSP-606M	OptoSigma
13	Optical bench carrer	CAA-65LSEE	OptoSigma
14	Optical table	HA-129-100L	OptoSigma

Table A.2: List of parts for Multi-frequency Illumination system. Duplicated items as shown in Table A.1 are omitted. Jigs are made in the workshop attached to the Institute of Scientific and Industrial Research, Osaka University.

Appendix B

Mathematical Study

In this appendices, we discuss in detail of our method that is omitted in the main text of Chapter 5 due to the readability.

B.1 Determining s from Hypothesized t using ToF Length

The back surface point $\mathbf{b}$ has two equality constraints among measurements and parameters. The first equality is that $\mathbf{b}$ is on the reference ray as

$$\mathbf{b} = \mathbf{r}_1 - s\mathbf{v}_3. \tag{B.1}$$

The other is that the optical length satisfies the ToF distortion model as

$$t + \nu \left|\mathbf{b} - \mathbf{f}\right| + \left|\mathbf{r}_1 - \mathbf{b}\right| = \tilde{d}. \tag{B.2}$$

The solution of the back surface point is the crossing point of the line (Eq. (B.1)) and quadratic surface (Eq. (B.2)) as shown in Fig. B.1.

Squaring Eq. (B.2) after substitution of $\mathbf{b}$ for Eq. (B.1) and transposition, we get

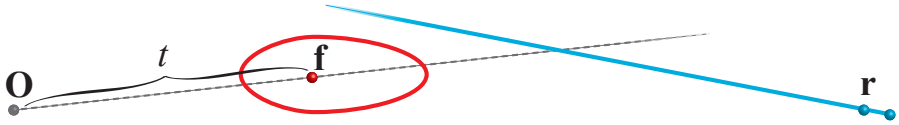
$$\nu^2 \left|\mathbf{r}_1 - s\mathbf{v}_3 - \mathbf{f}\right|^2 = \left(\tilde{d} - t - s\right)^2. \tag{B.3}$$

Organizing for s,

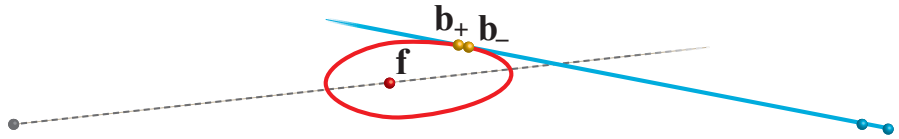
$$\begin{aligned} s^2(\nu^2 - 1) + 2s\left(\tilde{d} - t - \nu^2 \left(\mathbf{r}_1 - \mathbf{f}\right)^T \mathbf{v}_3\right) \\ +\nu^2 \left|\mathbf{r}_1 - \mathbf{f}\right|^2 - (\tilde{d} - t)^2 = 0. \end{aligned} \tag{B.4}$$

Finally, from the quadratic formula, we get

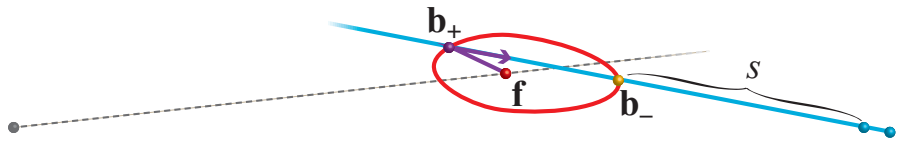
$$s = \frac{-h \pm \sqrt{h^2 - gi}}{g}, \tag{B.5}$$



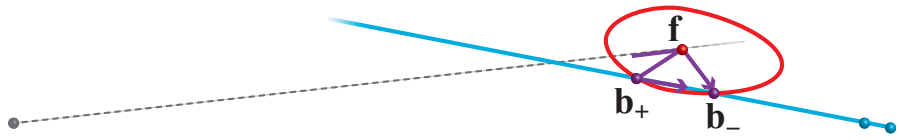
Case (a): Infeasible t. Quadratic equation has no solution.



Case (b): Two feasible solutions.



Case (c): Only smaller one is feasible.



Case (d): Both solutions (and t) is infeasible.

Figure B.1: Solution of the back surface point is a crossing point of reference ray (Eq. (B.1)) illustrated as blue line and quadratic surface (Eq. (B.2)) illustrated as red circle. (a) If t is too small, there is no intersection hence t is infeasible. (b) Two feasible solutions exist. (c) There are two solutions, but large one ($\mathbf{b}_+$) breaks total reflection condition (purple path). (d) If t is too large, both solutions break total reflection condition or they do not cross hence t is infeasible.

where g, h, and i are auxiliary variable such that

$$\begin{cases} g & = \nu^2 - 1 \\ h & = \tilde{d} - t - \nu^2 \left(\mathbf{r}_1 - \mathbf{f}\right)^T \mathbf{v}_3 \\ i & = \nu^2 \left|\mathbf{r}_1 - \mathbf{f}\right|^2 - \left(\tilde{d} - t\right)^2 . \end{cases} \tag{B.6}$$

There are some cases based on the relationship between the quadratic surface and reference ray. The first case is that the reference ray does not cross the quadratic surface as shown in Fig. B.1(a) because the hypothesized t is too small. In this case, $h^2 - gi < 0$ hence t is infeasible. The second case is that the reference ray is a tangent of the surface or similar situation as shown in (b). In this case, two solution $\mathbf{b}_-$ and $\mathbf{b}_+$ are both feasible. For the continuity of the following third case, we choose smaller s in our implementation. The third case, as shown in (c), is that one of the solution ($\mathbf{b}_+$) is infeasible because the light path from

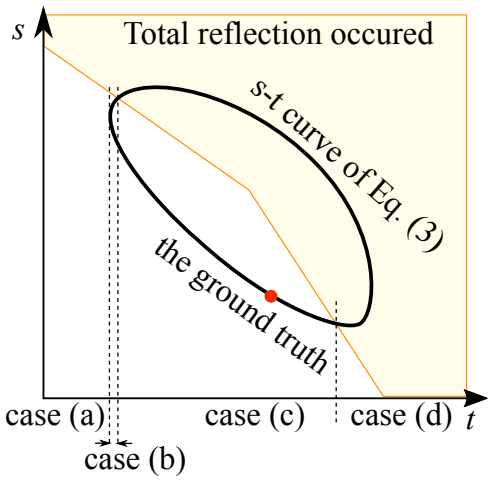


Figure B.2: Example of s-t solution space. Solution is on a quadratic curve of Eq. (B.3) and the red point is the ground-truth. Candidate light paths in the orange region does not satisfy total reflection condition.

$\mathbf{f} \rightarrow \mathbf{b}_+ \rightarrow \mathbf{r}_1$ does not satisfy total reflection condition

$$\mathbf{v_{in}}^T \mathbf{v_{out}} > \cos\theta_c, \tag{B.7}$$

where $\theta_c = \sin^{-1}(1/\nu)$ is the critical angle. In this case, only smaller s (*i.e.*, $\mathbf{b}_-$) is feasible. The last case is that the hypothesized depth t is too large as shown in (d). In this case, the smaller s is also infeasible because the light path $\mathbf{O} \rightarrow \mathbf{f} \rightarrow \mathbf{b}$ does not satisfy total reflection condition, where $\mathbf{O}$ is the camera position. Furthermore, it could be no solution ($h^2 - gi < 0$). Figure B.2 shows an example of s-t solution space. Orange region shows where light paths break total reflection condition, and we can see smaller s is the solution in most cases. This assumption breaks down only if the back refraction is near the total reflection limit, which can appear at the edge of the object.

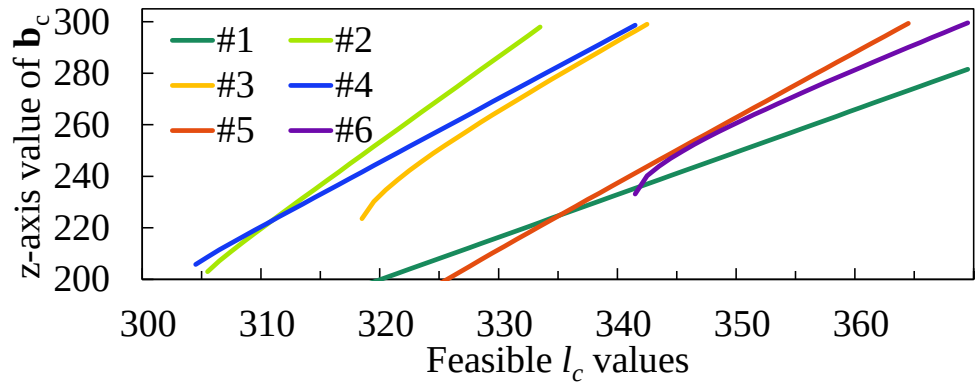


Figure B.3: The plot of f_c with respect to l_c. Samples are picked from various simulation objects. They are monotonically increasing in feasible l_c, and look like affine.

B.2 Normal Vector from Refractive Path

The definition of Snell's law in a cross product form is

$$\mathbf{v}_1 \times \mathbf{n} = \nu \mathbf{v}_2 \times \mathbf{n}, \tag{B.8}$$

where $\mathbf{v}_1$ is an incident ray, $\mathbf{v}_2$ is its refractive ray, and $\mathbf{n}$ is the normal of the surface. By transposition, we can obtain the surface normal direction as

$$\mathbf{n} \propto \nu \mathbf{v}_2 - \mathbf{v}_1. \tag{B.9}$$

By normalization, we can obtain the surface normal vector.

B.3 Convexity of l-subproblem

The value of z axis of the back point f_c can be represented as

$$f_c(l_c) = r_{1,c,z} - s(\hat{t}_c, l_c) v_{3,c,z}, \tag{B.10}$$

where $r_{1,c,z}$ and $v_{3,c,z}$ are z-axis value of $\mathbf{r}_1$ and $\mathbf{v}_3$ at pixel c, respectively. Using the function f_c, we can rewrite l-subproblem (Eq. 11 in the main text) as

$$\begin{aligned} \underset{\mathbf{l}}{\operatorname{argmin}} \sum_{c \in C} \left\| l_c - \tilde{d}(c) \right\|_2^2 + \\ \lambda_3'' \sum_{j,k \in N} \left\| f_j(l_j) - f_k(l_k) \right\|_H . \end{aligned} \tag{B.11}$$

If the function f_c is an affine function, the optimization problem can be rewritten into a variant of total variation denoising problem, which is known as a convex problem [140]. In our problem, f_c is not an affine function but we have observed that it has similar properties, which is

monotonic increase and small magnitude of second derivative. Figure B.3 shows example plots of the function f_c in feasible set of l_c, and they are similar to the affine function. If the light path candidate is near the total reflection condition, this affinity breaks down as shown in #3 and #6. In practice, we assume our objective function should have similar property; the global minimum can be obtained.

Reference

[1] C. Inoshita, "Shape from Scattering: Shape Estimation Based on Light Transport Analysis in Translucent Objects," *Ph. D. Thesis*, 2015.

[2] C. Inoshita, Y. Mukaigawa, Y. Matsushita, and Y. Yagi, "Surface Normal Decomposition: Photometric Stereo for Optically Thick Translucent Objects," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 346–359, 2014.

[3] C. Inoshita, S. Tagawa, M. A. Mannan, Y. Mukaigawa, and Y. Yagi, "Full-Dimensional Sampling and Analysis of BSSRDF," *IPSJ Transaction on Computer Vision and Applications (CVA)*, vol. 5, pp. 119–123, 2013.

[4] A. Kirmani, T. Hutchison, J. Davis, and R. Raskar, "Looking Around the Corner using Ultrafast Transient Imaging," *International Journal of Computer Vision (IJCV)*, vol. 95, no. 1, pp. 13–28, 2011.

[5] S. G. Narasimhan, M. Gupta, C. Donner, R. Ramamoorthi, S. K. Nayar, and H. W. Jensen, "Acquiring Scattering Properties of Participating Media by Dilution," *ACM Transactions on Graphics (ToG)*, vol. 25, no. 3, pp. 1003–1012, 2006.

[6] H. Li, S. Foo, K. E. Torrance, and S. H. Westin, "Automated Trhee-Axis Gonioreflectometer for Computer Graphics Applications," *Optical Engineering*, vol. 45, no. 4, p. 043605, 2006.

[7] G. Müller, G. H. Bendels, and R. Klein, "Rapid Synchronous Acquisition of Geometry and Appearance of Cultural Heritage Artefacts," in *Proceedings of the International Symposium on Vietual Reality, Archaeology and Cultural Heritage*, pp. 13–20, 2005.

[8] Y. Mukaigawa, K. Sumino, and Y. Yagi, "Multiplexed Illumination for Measuring BRDF using an Ellipsoidal Mirror and a Projector," in *Proceedings of the Asian Conference on Computer Vision (ACCV)*, pp. 246–257, 2007.

[9] W. Matusik, H. Pfister, M. Brand, and L. McMillan, "A Data-Driven Reflectance Model," *ACM Transactions on Graphics (ToG)*, vol. 22, no. 3, pp. 759–769, 2003.

[10] Y. Mukaigawa, K. Suzuki, and Y. Yagi, "Analysis of Subsurface Scattering under Generic Illumination," in *Proceedings of the IAPR International Conference on Pattern Recognition*, pp. 2452–2456, 2008.

[11] S. Tagawa, Y. Mukaigawa, and Y. Yagi, "8-D Reflectance Field for Computational Photography," in *Proceedings of the IAPR International Conference on Pattern Recognition*, pp. 2181–2185, 2012.

[12] S. K. Nayar, G. Krishnan, M. D. Grossberg, and R. Raskar, "Fast Separation of Direct and Global Components of a Scene using High Frequency Illumination," *ACM Transactions on Graphics (ToG)*, vol. 25, no. 3, pp. 935–944, 2006.

[13] B. Lamond, P. Peers, and P. Debevec, "Fast Image-Based Separation of Diffuse and Specular Reflections," in *Proceedings of ACM SIGGRAPH Sketches*, p. 74, 2007.

[14] Y. Mukaigawa, R. Raskar, and Y. Yagi, "Analysis of Light Transport in Scattering Media," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 153–160, 2010.

[15] Y. Mukaigawa, R. Raskar, and Y. Yagi, "Analysis of scattering light transport in translucent media," *IPSJ Transaction on Computer Vision and Applications (CVA)*, vol. 3, pp. 122–133, 2011.

[16] K. Tanaka, Y. Mukaigawa, Y. Matsushita, and Y. Yagi, "Descattering of Transmissive Observation using Parallel High-Frequency Illumination," in *Proceedings of the International Conference on Computational Photography (ICCP)*, pp. 96–103, IEEE, 2013.

[17] K. Tanaka, Y. Mukaigawa, H. Kubo, Y. Matsushita, and Y. Yagi, "Recovering Inner Slices of Translucent Objects by Multi-Frequency Illumination," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5464–5472, 2015.

[18] S. Achar, S. T. Nuske, and S. G. Narasimhan, "Compensating for Motion During Direct-Global Separation," in *Proceedings of the International Conference on Computer Vision (ICCV)*, pp. 1481–1488, 2013.

[19] S. Achar and S. G. Narasimhan, "Multi Focus Structured Light for Recovering Scene Shape and Global Illumination," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 205–219, 2014.

[20] D. Reddy, R. Ramamoorthi, and B. Curless, "Frequency-Space Decomposition and Acquisition of Light Transport under Spatially Varying Illumination," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 596–610, 2012.

[21] M. Gupta, Y. Tian, S. G. Narasimhan, and L. Zhang, "A Combined Theory of Defocused Illumination and Global Light Transport," *International Journal of Computer Vision (IJCV)*, vol. 98, no. 2, pp. 146–167, 2011.

[22] M. O'Toole, R. Raskar, and K. N. Kutulakos, "Primal-Dual Coding to Probe Light Transport," *ACM Transactions on Graphics (ToG)*, vol. 31, no. 4, pp. 39:1–39:11, 2012.

[23] M. O'Toole, J. Mather, and K. N. Kutulakos, "3D Shape and Indirect Appearance by Structured Light Transport," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3246–3253, IEEE, 2014.

[24] M. O'Toole, S. Achar, S. G. Narasimhan, and K. N. Kutulakos, "Homogeneous Codes for Energy-Efficient Illumination and Imaging," *ACM Transactions on Graphics (ToG)*, vol. 34, no. 4, pp. 35:1–35:13, 2015.

[25] M. Levoy, B. Chen, V. Vaish, M. Horowitz, I. McDowall, and M. Bolas, "Synthetic Aperture Confocal Imaging," *ACM Transactions on Graphics (ToG)*, vol. 23, no. 3, pp. 825–834, 2004.

[26] M. Levoy and P. Hanrahan, "Light Field Rendering," in *Proceedings of ACM SIGGRAPH*, pp. 31–42, 1996.

[27] T. R. Corle and G. S. Kino, eds., *Confocal Scanning Optical Microscopy and Related Imaging Systems*. Academic Press, 1996.

[28] C. Fuchs, M. Heinz, M. Levoy, H.-P. Seidel, and H. P. A. Lensch, "Combining Confocal Imaging and Descattering," *Computer Graphics Forum*, vol. 27, no. 4, pp. 1245–1253, 2008.

[29] S. G. Narasimhan, S. K. Nayar, and S. Koppal, "Structured Light in Scattering Media," in *Proceedings of the International Conference on Computer Vision (ICCV)*, pp. 420–427, IEEE, 2005.

[30] J. Gu, S. K. Nayar, E. Grinspun, P. Belhumeur, and R. Ramamoorthi, "Compressive Structured Light for Recovering Inhomogeneous Participating Media," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 35, no. 3, pp. 555–567, 2013.

[31] I. Gkioulekas, A. Levin, F. Durand, and T. Zickler, "Micron-Scale Light Transport Decomposition using Interferometry," *ACM Transactions on Graphics (ToG)*, vol. 34, no. 4, pp. 37:1–37:14, 2015.

[32] N. Abramson, "Light-in-Flight Recording by Holography," *Optics Letters*, vol. 3, no. 4, pp. 121–123, 1978.

[33] T. Kakue, K. Tosa, J. Yuasa, T. Tahara, Y. Awatsuji, K. Nishio, S. Ura, and T. Kubota, "Digital Light-in-Flight Recording by Holography by Use of a Femtosecond Pulsed Laser," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 18, no. 1, pp. 479–485, 2012.

[34] A. Velten, R. Raskar, D. Wu, A. Jarabo, B. Masia, C. Barsi, C. Joshi, E. Lawson, M. Bawendi, and D. Gutierrez, "Femto-Photography: Capturing and Visualizing the Propagation of Light," *ACM Transactions on Graphics (ToG)*, vol. 32, no. 4, pp. 44:1–44:8, 2013.

[35] A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. G. Bawendi, and R. Raskar, "Recovering Three-Dimensional Shape Around a Corner using Ultrafast Time-of-Flight Imaging.," *Nature communications*, vol. 3, pp. 745:1–745:8, 2012.

[36] F. Heide, M. B. Hullin, J. Gregson, and W. Heidrich, "Low-Budget Transient Imaging using Photonic Mixer Devices," *ACM Transactions on Graphics (ToG)*, vol. 32, no. 4, pp. 45:1–45:10, 2013.

[37] J. Lin, Y. Liu, M. B. Hullin, and Q. Dai, "Fourier Analysis on Transient Imaging with a Multifrequency Time-of-Flight Camera," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3230–3237, IEEE, 2014.

[38] C. Peters, J. Klein, M. B. Hullin, and R. Klein, "Solving Trigonometric Moment Problems for Fast Transient Imaging," *Proceedings of ACM SIGGRAPH Asia*, vol. 34, no. 6, pp. 220:1–220:11, 2015.

[39] A. Kadambi, R. Whyte, A. Bhandari, L. Streeter, C. Barsi, A. Dorrington, and R. Raskar, "Coded Time of Flight Cameras: Sparse Deconvolution to Address Multipath Interference and Recover Time Profiles," *ACM Transactions on Graphics (ToG)*, vol. 32, no. 6, pp. 167:1–167:10, 2013.

[40] M. O'Toole, F. Heide, L. Xiao, M. B. Hullin, W. Heidrich, and K. N. Kutulakos, "Temporal Frequency Probing for 5D Transient Analysis of Global Light Transport," *ACM Transactions on Graphics (ToG)*, vol. 33, no. 4, pp. 87:1–87:11, 2014.

[41] S. Fuchs, "Multipath Interference Compensation in Time-of-Flight Camera Images," in *Proceedings of the IAPR International Conference on Pattern Recognition*, pp. 3583–3586, IEEE, 2010.

[42] A. A. Dorrington, J. P. Godbaz, M. J. Cree, A. D. Payne, and L. V. Streeter, "Separating True Range Measurements from Multi-Path and Scattering Interference in Commercial Range Cameras," in *SPIE 7864, Three-Dimensional Imaging, Interaction, and Measurement*, p. 786404, 2011.

[43] J. P. Godbaz, M. J. Cree, and A. A. Dorrington, "Closed-Form Inverses for the Mixed Pixel/Multipath Interference Problem in AMCW Lider," in *SPIE 8296, Computational Imaging X*, p. 909778, 2012.

[44] D. Jimenez, D. Pizarro, M. Mazo, and S. Palazuelos, "Modelling and Correction of Multipath Interference in Time of Flight Cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 893–900, IEEE, 2012.

[45] F. Heide, L. Xiao, A. Kolb, M. B. Hullin, and W. Heidrich, "Imaging in Scattering Media using Correlation Image Sensors and Sparse Convolutional Coding.," *Optics express*, vol. 22, no. 21, pp. 26338–26350, 2014.

[46] A. Kirmani, A. Benedetti, and P. A. Chou, "SPUMIC: Simultaneous Phase Unwrapping and Multipath Interference Cancellation in Time-of-Flight Cameras using Spectral Methods," in *IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, IEEE, 2013.

[47] A. Bhandari, M. Feigin, S. Izadi, C. Rhemann, M. Schmidt, and R. Raskar, "Resolving Multipath Interference in Kinect: an Inverse Problem Approach," in *IEEE SENSORS*, pp. 614–617, IEEE, 2014.

[48] D. Freedman, E. Krupka, Y. Smolin, I. Leichter, and M. Schmidt, "SRA: Fast Removal of General Multipath for ToF Sensors," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 234–249, 2014.

[49] H. Qiao, J. Lin, Y. Liu, M. B. Hullin, and Q. Dai, "Resolving Transient Time Profile in ToF Imaging via Log-Sum Sparse Regularization.," *Optics letters*, vol. 40, no. 6, pp. 918–21, 2015.

[50] S. Lee and H. Shim, "Skewed Stereo Time-of-Flight Camera for Translucent Object Imaging," *Image and Vision Computing*, vol. 43, no. C, pp. 27–38, 2015.

[51] N. Naik, A. Kadambi, C. Rhemann, S. Izadi, R. Raskar, and S. Bing Kang, "A Light Transport Model for Mitigating Multipath Interference in Time-of-Flight Sensors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 73–81, 2015.

[52] A. Kadambi, J. Schiel, and R. Raskar, "Macroscopic Interferometry: Rethinking Depth Estimation with Frequency-Domain Time-of-Flight," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 893–902, 2016.

[53] H. Shim and S. Lee, "Recovering Translucent Object using a Single Time-of-Flight Depth Camera," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 5, pp. 841–854, 2015.

[54] K. Tanaka, Y. Mukaigawa, H. Kubo, Y. Matsushita, and Y. Yagi, "Recovering Transparent Shape from Time-of-Flight Distortion," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4387–4395, 2016.

[55] N. Naik, S. Zhao, A. Velten, R. Raskar, and K. Bala, "Single View Reflectance Capture using Multiplexed Scattering and Time-of-Flight Imaging," *ACM Transactions on Graphics (ToG)*, vol. 30, no. 6, pp. 171:1–171:10, 2011.

[56] F. Heide, L. Xiao, W. Heidrich, and M. B. Hullin, "Diffuse Mirrors: 3D Reconstruction from Diffuse Indirect Illumination using Inexpensive Time-of-Flight Sensors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3222–3229, 2014.

[57] A. Kadambi, H. Zhao, B. Shi, and R. Raskar, "Occluded Imaging with Time-of-Flight Sensors," *ACM Transactions on Graphics (ToG)*, vol. 35, no. 2, pp. 15:1–15:12, 2016.

[58] D. Wu, A. Velten, M. O'Toole, B. Masia, A. Agrawal, Q. Dai, and R. Raskar, "Decomposing Global Light Transport using Time of Flight Imaging," *International Journal of Computer Vision (IJCV)*, vol. 107, no. 2, pp. 123–138, 2014.

[59] M. Gupta, S. K. Nayar, M. B. Hullin, and J. Martin, "Phasor Imaging: a Generalization of Correlation-Based Time-of-Flight Imaging," *ACM Transactions on Graphics (ToG)*, vol. 34, no. 5, pp. 156:1–156:18, 2015.

[60] F. Heide, W. Heidrich, M. Hullin, and G. Wetzstein, "Doppler Time-of-Flight Imaging," *ACM Transactions on Graphics (ToG)*, vol. 34, no. 4, pp. 36:1–36:11, 2015.

[61] S. Shrestha, F. Heide, W. Heidrich, and G. Wetzstein, "Computational Imaging with Multi-Camera Time-of-Flight Systems," *ACM Transactions on Graphics (ToG)*, vol. 35, no. 4, pp. 33:1–33:11, 2016.

[62] R. Tadano, A. Kumar Pediredla, and A. Veeraraghavan, "Depth Selective Camera: a Direct, On-Chip, Programmable Technique for Depth Selectivity in Photography," in *Proceedings of the International Conference on Computer Vision (ICCV)*, pp. 3595–3603, 2015.

[63] D. Boas and M. A. Franceschini, "Near Infrared Imaging," *Scholarpedia*, vol. 4, no. 4, p. 6997, 2009.

[64] G. S. Settles, *Schlieren and Shadowgraph Techniques: Visualizing Phenomena in Transparent Media*. Springer, 2001.

[65] D. Brogioli, A. Vailati, and M. Giglio, "A Schlieren Method for Ultra-Low Angle Light Scattering Measurements," *EPL (Europhysics Letters)*, vol. 63, no. 2, pp. 220–226, 2003.

[66] G. Wetzstein, D. Roodnick, W. Heidrich, and R. Raskar, "Refractive Shape from Light Field Distortion," in *Proceedings of the International Conference on Computer Vision (ICCV)*, pp. 1180–1186, IEEE, 2011.

[67] G. D. Gilbert and J. C. Pernicka, "Improvement of Underwater Visibility by Reduction of Backscatter with a Circular Polarization Technique," *Applied Optics*, vol. 6, no. 4, pp. 741–746, 1967.

[68] T. Treibitz and Y. Y. Schechner, "Active Polarization Descattering," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 31, no. 3, pp. 385–399, 2009.

[69] J. Kim, D. Lanman, Y. Mukaigawa, and R. Raskar, "Descattering Transmission via Angular Filtering," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 86–99, 2010.

[70] K. He, J. Sun, and X. Tang, "Single Image Haze Removal using Dark Channel Prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 33, no. 12, pp. 2341–2353, 2011.

[71] K. Nishino, L. Kratz, and S. Lombardi, "Baysian Defogging," *International Journal of Computer Vision (IJCV)*, vol. 98, no. 3, pp. 263–278, 2012.

[72] R. Fattal, "Dehazing using color-lines," *ACM Transactions on Graphics (ToG)*, vol. 34, no. 1, pp. 13:1–13:14, 2014.

[73] D. Berman, T. Treibitz, and S. Avidan, "Non-Local Image Dehazing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1674–1682, 2016.

[74] R. Szeliski, S. Avidan, and P. Anandan, "Layer Extraction from Multiple Images Containing Reflections and Transparency," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 246–253, IEEE Computer Society, 2000.

[75] B. Sarel and M. Irani, "Separating Transparent Layers through Layer Information Exchange," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 328–341, 2004.

[76] Y. Li and M. S. Brown, "Single Image Layer Separation using Relative Smoothness," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2752–2759, 2014.

[77] T. Poter and T. Duff, "Compositing Digital Images," *Computer Graphics*, vol. 18, no. 3, pp. 253–259, 1984.

[78] C. Donner and H. W. Jensen, "Light Diffusion in Multi-Layered Translucent Materials," *ACM Transactions on Graphics (ToG)*, vol. 24, no. 3, pp. 1032–1039, 2005.

[79] H. W. Jensen, S. R. Marschner, M. Levoy, and P. Hanrahan, "A Practical Model for Subsurface Light Transport," in *Proceedings of ACM SIGGRAPH*, pp. 511–518, ACM Press, 2001.

[80] P. Kubelka and F. Munk, "Ein Beitrag Zur Optik Der Farbanstriche," *Zeitschrift fr Technische Physik*, vol. 12, pp. 593–601, 1931.

[81] E. D'Eon, G. Irving, E. D'Eon, and G. Irving, "A Quantized-Diffusion Model for Rendering Translucent Materials," *ACM Transactions on Graphics (ToG)*, vol. 30, no. 4, pp. 56:1–56:14, 2011.

[82] "Green Function," in *Encyclopedia of Mathematics* (M. Hazewinkel, ed.), Springer, 2001.

[83] W. Jakob, E. D'Eon, O. Jakob, and S. Marschner, "A Comprehensive Framework for Rendering Layered Materials," *ACM Transactions on Graphics (ToG)*, vol. 33, no. 4, pp. 118:1–118:14, 2014.

[84] D. Gavrilov, E. Maeva, O. Grube, X. Maldague, and R. Maev, "Infrared Methods in Noninvasive Inspection of Artwork," in *International Conference on NDT of Art*, pp. 1–5, 2008.

[85] J. Dik, K. Janssens, G. Van der Snickt, L. van der Loeff, K. Rickers, and M. Cotte, "Visualization of a Lost Painting by Vincent Van Gogh using Synchrotron Radiation Based X-Ray Fluorescence Elemental Mapping.," *Analytical chemistry*, vol. 80, no. 16, pp. 6436–42, 2008.

[86] A. J. L. Adam, P. C. M. Planken, S. Meloni, and J. Dik, "Terahertz Imaging of Hidden Paint Layers on Canvas," *Optics Express*, vol. 17, no. 5, pp. 3407–3416, 2009.

[87] A. R. Sanchez, B. Heshmat, A. Aghasi, M. Zhang, S. Naqvi, J. Romberg, and R. Raskar, "Terahertz Time-gated Spectroscopic Imaging for Content Extraction through Layered Structures," *Nature Communications*, vol. 7, no. 12665, pp. 1–7, 2016.

[88] D. Huang, E. Swanson, C. Lin, J. Schuman, W. Stinson, W. Chang, M. Hee, T. Flotte, K. Gregory, C. Puliafito, and A. Et, "Optical Coherence Tomography," *Science*, vol. 254, no. 5035, pp. 1178–1181, 1991.

[89] A. F. Fercher, "Optical Coherence Tomography.," *Journal of biomedical optics*, vol. 1, no. 2, pp. 157–73, 1996.

[90] M. G. L. Gustafsson, L. Shao, P. M. Carlton, C. J. R. Wang, I. N. Golubovskaya, W. Z. Cande, D. A. Agard, and J. W. Sedat, "Three-Dimensional Resolution Doubling in Wide-Field Fluorescence Microscopy by Structured Illumination.," *Biophysical journal*, vol. 94, no. 12, pp. 4957–70, 2008.

[91] P. Kner, B. B. Chhun, E. R. Griffis, L. Winoto, and M. G. L. Gustafsson, "Super-Resolution Video Microscopy of Live Cells by Structured Illumination.," *Nature methods*, vol. 6, no. 5, pp. 339–42, 2009.

[92] O. D. Therrien, B. Aubé, S. Pagès, P. D. Koninck, and D. Côté, "Wide-Field Multiphoton Imaging of Cellular Dynamics in Thick Tissue by Temporal Focusing and Patterned Illumination.," *Biomedical optics express*, vol. 2, no. 3, pp. 696–704, 2011.

[93] N. J. W. Morris and K. N. Kutulakos, "Reconstructing the Surface of Inhomogeneous Transparent Scenes by Scatter-Trace Photography," in *Proceedings of the International Conference on Computer Vision (ICCV)*, pp. 1–8, 2007.

[94] T. Hawkins, P. Einarsson, and P. Debevec, "Acquisition of Time-Varying Participating Media," *ACM Transactions on Graphics (ToG)*, vol. 24, no. 3, pp. 812–815, 2005.

[95] I. Ihrke and M. Magnor, "Adaptive Grid Optical Tomography," *Graphical Models*, vol. 68, no. 5-6, pp. 484–495, 2006.

[96] C. Inoshita, Y. Mukaigawa, Y. Matsushita, and Y. Yagi, "Shape from Single Scattering for Translucent Objects," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 371–384, 2012.

[97] H. Murase, "Surface Shape Reconstruction of an Undulating Transparent Object," in *Proceedings of the International Conference on Computer Vision (ICCV)*, pp. 313–317, IEEE Computer Society Press, 1990.

[98] N. J. W. Morris and K. N. Kutulakos, "Dynamic Refraction Stereo.," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 33, no. 8, pp. 1518–1531, 2011.

[99] M. Alterman, Y. Y. Schechner, and Y. Swirski, "Triangulation in Random Refractive Distortions," in *Proceedings of the International Conference on Computational Photography (ICCP)*, pp. 1–10, IEEE, 2013.

[100] Y. Tian and S. G. Narasimhan, "Seeing through Water: Image Restoration using Model-Based Tracking," in *Proceedings of the International Conference on Computer Vision (ICCV)*, pp. 2303–2310, IEEE, 2009.

[101] Y. Tian and S. G. Narasimhan, "A Globally Optimal Data-Driven Approach for Image Distortion Estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1277–1284, IEEE, 2010.

[102] K. Kutulakos and E. Steger, "A Theory of Refractive and Specular 3D Shape by Light-Path Triangulation," *International Journal of Computer Vision (IJCV)*, vol. 76, no. 1, pp. 13–29, 2008.

[103] S. Yeung, T. Wu, C. Tang, T. F. Chan, and S. Osher, "Adequate Reconstruction of Transparent Objects on a Shoestring Budget," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2513–2520, IEEE, 2011.

[104] K. Han, K.-Y. K. Wong, and M. Liu, "A Fixed Viewpoint Approach for Dense Reconstruction of Transparent Objects," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4001–4008, 2015.

[105] I. Ihrke, B. Goidluecke, and M. Magnor, "Reconstructing the Geometry of Flowing Water," in *Proceedings of the International Conference on Computer Vision (ICCV)*, pp. 1055–1060, IEEE, 2005.

[106] C. Ma, X. Lin, J. Suo, Q. Dai, and G. Wetzstein, "Transparent Object Reconstruction via Coded Transport of Intensity," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3238–3245, IEEE, 2014.

[107] D. Miyazaki and K. Ikeuchi, "Inverse Polarization Raytracing: Estimating Surface Shapes of Transparent Objects," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 910–917, IEEE, 2005.

[108] G. Eren, O. Aubreton, F. Meriaudeau, L. A. Sanchez Secades, D. Fofi, A. T. Naskali, F. Truchetet, and A. Ercil, "Scanning from Heating: 3D Shape Estimation of Transparent Objects from Local Surface Heating.," *Optics express*, vol. 17, no. 14, pp. 11457–68, 2009.

[109] B. Caputo, E. Hayman, and P. Mallikarjuna, "Class-Specific Material Categorisation," in *Proceedings of the International Conference on Computer Vision (ICCV)*, pp. 1597–1604, IEEE, 2005.

[110] M. Varma and A. Zisserman, "A Statistical Approach to Material Classification using Image Patch Exemplars," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 31, no. 11, pp. 2032–2047, 2009.

[111] C. Liu, L. Sharan, E. H. Adelson, and R. Rosenholtz, "Exploring Features in a Bayesian Framework for Material Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 239–246, IEEE, 2010.

[112] G. Schwartz and K. Nishino, "Visual Material Traits: Recognizing Per-Pixel Material Context," *Proceedings of Color and Photometry in Computer vision workshop in conjunction with the IEEE International Conference on Computer Vision*, pp. 883–890, 2013.

[113] G. Schwartz and K. Nishino, "Automatically Discovering Local Visual Material Attributes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3565–3573, 2015.

[114] S. Bell, P. Upchurch, N. Snavely, and K. Bala, "Material Recognition in the Wild with the Materials in Context Database," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3479–3487, 2015.

[115] H. Zheng, L. Fang, M. JI, M. Strese, Y. Oezer, and E. Steinbach, "Deep Learning for Surface Material Classification using Haptic and Visual Information," *IEEE Transactions on Multimedia*, vol. 18, no. 12, pp. 2407–2416, 2016.

[116] T.-C. Wang, J.-Y. Zhu, E. Hiroaki, M. Chandraker, A. A. Efros, and R. Ramamoorthi, "A 4D Light-Field Dataset and CNN Architectures for Material Recognition," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 121–138, Springer International Publishing, 2016.

[117] H. Zhang, K. Dana, and K. Nishino, "Reflectance Hashing for Material Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3071–3080, 2015.

[118] C. Liu and J. Gu, "Discriminative Illumination: Per-Pixel Classification of Raw Materials Based on Optimal Projections of Spectral BRDF," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 36, no. 1, pp. 86–98, 2014.

[119] M. A. Mannan, D. Das, Y. Kobayashi, and Y. Kuno, "Object Material Classification by Surface Reflection Analysis with a Time-of-Flight Range Sensor," in *Advances in Visual Computing*, pp. 439–448, Springer Berlin Heidelberg, 2010.

[120] M. Sato, S. Yoshida, A. Olwal, B. Shi, A. Hiyama, T. Tanikawa, M. Hirose, and R. Raskar, "SpecTrans: Versatile Material Classification for Interaction with Textureless, Specular and Transparent Surfaces," in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 2191–2200, ACM Press, 2015.

[121] A. Davis, K. L. Bouman, J. G. Chen, M. Rubinstein, F. Durand, and W. T. Freeman, "Visual Vibrometry: Estimating Material Properties from Small Motion in Video," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5335–5343, 2015.

[122] P. Saponaro, S. Sorensen, A. Kolagunda, and C. Kambhamettu, "Material Classification with Thermal Imagery," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4649–4656, 2015.

[123] S. Su, F. Heide, R. Swanson, J. Klein, C. Callenberg, M. Hullin, and W. Heidrich, "Material Classification using Raw Time-of-Flight Measurements," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3503–3511, 2016.

[124] N. Miura and Y. Sato, "Deblurring Vein Images and Removing Skin Wrinkle Patterns by using Tri-Band Illumination," in *Proceedings of the Asian Conference on Computer Vision (ACCV)*, pp. 336–349, 2012.

[125] K. Shimizu, K. Tochio, and Y. Kato, "Improvement of Transcutaneous Fluorescent Images with a Depth-Dependent Point-Spread Function," *Applied Optics*, vol. 44, no. 11, pp. 2154–2161, 2005.

[126] S. G. Narasimhan and S. K. Nayar, "Shedding Light on the Weather," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 665–672, 2003.

[127] S. Chandrasekhar, *Radiative Transfer*. Dover Publications, Inc., 1960.

[128] A. Ishimaru, *Wave Propagation and Scattering in Random Media*. IEEE Press, 1997.

[129] P. Hanrahan and W. Krueger, "Reflection from Layered Surfaces Due to Subsurface Scattering," in *Proceedings of ACM SIGGRAPH*, pp. 165–174, ACM Press, 1993.

[130] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1994.

[131] P. Pérez, M. Gangnet, and A. Blake, "Poisson Image Editing," *ACM Transactions on Graphics (ToG)*, vol. 22, no. 3, pp. 313–318, 2003.

[132] N. Kawai and N. Yokoya, "Image Inpainting Considering Symmetric Patterns," in *Proceedings of the IAPR International Conference on Pattern Recognition*, pp. 2744–2747, 2012.

[133] P. J. Huber, "Robust Regression: Asymptotics, Conjectures and Monte Carlo," *The Annals of Statistics*, vol. 1, no. 5, pp. 799–821, 1973.

[134] C. Zhu, R. H. Byrd, and J. Nocedal, "L-BFGS-B:, Algorithm 778: L-BFGS-B, Fortran Routines for Large Scale Bound Constrained Optimization," *ACM Transactions on Mathematical Software*, vol. 23, no. 4, pp. 550–560, 1997.

[135] A. Buades, B. Coll, and J.-M. Morel, "A Non-Local Algorithm for Image Denoising," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 60–65, IEEE, 2005.

[136] S. Inokuchi, K. Sato, and F. Matsuda, "Range Imaging System for 3-D Object Recognition," in *Proceedings of the IAPR International Conference on Pattern Recognition*, pp. 806–808, 1984.

[137] P. Besl and H. McKay, "A Method for Registration of 3-D Shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 14, no. 2, pp. 239–256, 1992.

[138] Z. Xu, T. Perry, and G. Hills, "Method and system for multi-phase dynamic calibration of three-dimensional (3d) sensors in a time-of-flight system," *US Patent 8587771 B2*, 2013.

[139] J. P. Godbaz, A. Dorrington, and M. J. Cree, “Understanding and Ameliorating Mixed Pixels and Multipath Interference in AMCW Lidar,” in *TOF Range-Imaging Cameras.*, pp. 91–116, 2013.

[140] M. A. Little and N. S. Jones, “Sparse Bayesian Step-Filtering for High-Throughput Analysis of Molecular Machine Dynamics,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4162–4165, 2010.