

Shape from Multi-Bounce Light Paths

*Submitted in partial fulfillment of the requirements for
the degree of
Doctor of Philosophy
in
Department of Electrical and Computer Engineering*

Chia-Yin Tsai

B.S., Electrical Engineering, National Taiwan University
M.S., Electrical and Computer Engineering, Carnegie Mellon University

Carnegie Mellon University
Pittsburgh, PA

May 2019

© 2019 Chia-Yin Tsai.
All rights reserved.

Acknowledgments

This thesis is not possible without help and support from mentors, collaborators, friends, and family. The professional and personal improvements are beyond my imagination before starting the program. There are so many things that I could have done better, but looking back, I am immensely grateful for all the good and bad along this journey. This dissertation was supported in part by an NSF grant CCF-1117939, DARPA REVEAL Co-operative Agreement HR0011-16-2-0021, a gift from Adobe Research, CIT Deans Tuition Fellowship, Bertucci Graduate Fellowship, and Google PhD Fellowship.

I cannot thank my adviser Aswin Sankaranarayanan enough. It has been a privilege to be one of his first students. One thing I most appreciate is how he is (or appear to be) calm at all times. Nothing is too big a deal and just put down actionable items and try to tackle one thing at a time. The version 0.0 of this thesis is the Qualcomm innovation fellowship proposal jointly written with Hui Zhuo. Then version 0.1 is a more extended version for the Google PhD Fellowship application. Throughout the years, we put together block by block and the end result is this thesis. Thank you Aswin, for your vision so that we are collecting the correct blocks and for your being always available. He not only stands in my shoe but also consider things that I may have overlooked.

The image science lab labmates are my wonderful companions. Thanks to Vishwanath Saragadam to be my office mate and spend so much time talking sense and non-sense with me. Thank you Jian Wang for all the laughter and joy (and somehow finished our physics-based vision class project). Thank you Hui Zhuo for the version 0.0 of this thesis and the ride during ICCP 2016 submission season. Thank you Rick Chang for the debate over the best pineapple cake (and I will always be the winner). Somehow, after Yi Hua, Anqi Yang, Byeongjoo Ahn, Kuldeep Kulkarni, Vijay Rengarajan, Michael De Zeeuw, and Wei-Yu Chen join the lab, I am one of the "senior" PhDs. Thank you all for the support and time we spent together.

I would like to thank my committee members: Jelena Kovačević, Srinivasa G. Narasimhan, Kiriakos N. Kutulakos, and Ioannis Gkioulekas. I appreciate the insightful inputs and comments to make this thesis more complete. I would like to thank Jelena for spending time with me to go through the details of my research. Having her on my committee helps me to think about how this thesis can be applicable to a wider audience other than only people in computational photography. I learn so much from Srinivas how to tell a research story. The pep talk for me to remake the CVPR spotlight presentation is one of the coolest moments in my PhD program. I would like to thank Srinivas for his encouraging words and help. Kyros' dedication to work is beyond description. I will never forget how he went through my CVPR spotlight presentation word by word the day before the presentation. The video is pre-recorded

so putting sentences that can fit into the timing is pretty challenging. Even though the process is hard, the end result is totally worth it. It has been a wonderful year working closely with Yannis. The Monday 8:30 pm meetings are the best!! I would like to thank Yannis for his help and guidance through the project. The surface optimization chapter pushes this thesis to another level and the CVPR submission as an end of my PhD journey is just perfect.

I collaborated with Ashok Veeraraghavan on several projects. I have the chance to visit Rice University for ICCP and for data collection. Thank you, Ashok, for hospitality and experience sharing. It is wonderful to interact with him and members in his lab: Adithya Pediredla, George Chen, Vivek Boominathan, Souptik Barua, Ewa Nowara, and Jason Holloway. Thank you friends for my stay in Houston.

I am fortunate to work as an intern with the Chauffeur team at Google X (now Waymo). Thanks for the wonderful summer with members of the magic team: Abhijit Ogale, Wan-yen Lo, Andreas Wendel, Yuandong Tian, and Dave Ferguson. My mentor, Abhijit provides immense support during my internship and also as a mentor for the Google PhD fellowship program. It's during this internship that I see how a PhD training can help to put ideas into industry. Special thanks to Angjoo Kanazawa for all the discussion about academic life during the internship and also for the many update emails and meet-ups in conferences in the following years.

My internship with Oculus Research Pittsburgh (now Facebook Reality Lab Pittsburgh) helps me understand the immense gap between research ideas and practical uses. Thanks for members across Oculus Research: Iain Matthews, Chenglei Wu, Yaser Sheik, Alexander Fix, Lingling Tao, and Robert Cavin for valuable inputs for my project. I appreciate my mentors Iain and Chenglei's trust and support during the project to let me explore and develop both theoretical and experimental results. I am thankful for the friendship with Shou-I Yu, Shih-En Wei, Aayush Bansal, Junko Saragih, Ying Yang, Sahana Vijai, Alex Poms, and Dave Greenwood.

It is impossible to name everyone that help me throughout this journey, but I will try. I am also sure I will forget to include some important ones, but you know you are in my heart. My roommates, Rushane, Fish, Yu-Hsin, Hsueh-Lin, Chih-Ying, and Christine, made Pittsburgh home for me. Yen-Chi, De-An, Shihyun, Yi-Ju, Chen-Hsuan, Mini, Suren, Brian, Alankar, Satwik, Lianyan, Shumian, Chao made CMU life awesome. Coaches and friends in Pittsburgh Fit bring my physical health back: Amanda, Jodi, Jamie, Eddie, Erika, and Dave. Special thanks to Jenny, Grace, Chendi, Phoebe, Lingshu, Bob, Rachael, Yoonah, Liz, Yohan, Eugene, Aurora, Hush, Rachael, Nadia, William, Kim, Joey, Pei, Hope, Emily, Ellen, and Tanya, for spiritual shepherding. Friends back home in Taiwan show so much support for me, Chia-Chun, Ginny, Canon, Jocelyn, Yi-Ching, Jei Lin, Cheng Tsao, Chrissy, Kate, and Roxanne. Finally, my loving parents, Run-Yong and Chin-Hui, sister, Chun-Han, brother, Kevin, and Aunt, Hsuan-Ru, have always been my strongest supporters.

Let me end with a classic excerpt from Chinese textbook. The people that I have to express gratitude are too many, so I thank God.

Abstract

Classical shape estimation techniques often have simplified but unrealistic assumptions, such as light interacting with the scene with only single-bounce light paths and the scene being Lambertian. However, a real scene often interacts with light in significantly more complex ways; multi-bounce light paths are ubiquitous and real-world materials are typically non-Lambertian. In such cases, shape estimation is not just challenging, but also beyond the reach of commonly used techniques.

We propose a shape estimation framework that deals with multi-bounce light paths. The proposed framework uses light paths, as opposed to images, as the primitive for shape estimation. The core of our idea is that we can trace the optical journey of a collection of photons as light paths with multiple bounces, where each bounce is an instance of a light-object interaction. These interactions can often be explained by simple physical laws that govern how properties of light rays change; for example, a mirror simply changes the orientation of light rays while preserving their radiance; a diffuse wall scatters light in all directions and changes the light ray's radiance. Shape estimation problem now reduces to identifying underlying physical phenomena related to each bounce in the light path.

Our proposed shape recovery framework is particularly effective for scenes that interact with light in complex ways. We explore three such scenarios. First, we characterize how information pertaining to the shape of a transparent object is encoded in the deflection of light rays and use it to recover the shape of transparent objects. Second, we characterize how the path length of two-bounce light paths are sufficient for concave shape recovery. Finally, we show a specialized scenario where the object of interest can only be imaged through multi-bounce light paths. In all three scenarios, we show that by using physical properties related to each bounce of light paths, we can find constraints on the geometry of the scene, this leads to accurate shape estimation algorithms. The techniques developed in this thesis are applicable to a wide range of research fields. The image formation discussed in this thesis shares a lot of similarities in other fields, such as medical imaging, acoustic imaging, and wifi localization. We hope this thesis can inspire more researchers to deal with multi-bounce effects in different research fields.

Contents

1	Introduction	1
1.1	Related work	2
1.2	Contributions	7
2	Transparent object reconstruction	
	— Shape from deflection of two-bounce light paths	10
2.1	Problem statement	11
2.2	Shape from light-ray correspondences	11
2.3	Experiments	15
2.4	Conclusion	15
3	Concave object reconstruction	
	— Shape from time-of-flight of two-bounce light paths	19
3.1	Related work	20
3.2	Problem statement	20
3.3	Unique shape recovery analysis with two-bounce light paths	21
3.4	Experiments	26
3.5	Conclusion	28
4	Non-line-of-sight imaging	
	— Shape from shortest three-bounce light paths	33
4.1	Related work	35
4.2	Problem statement	37
4.3	Acquisition of first-returning photons	38
4.4	Geometry of first-returning photons	40
4.5	Shape from first-returning photons	51
4.6	Experiments	56
	4.6.1 Simulation results	56
	4.6.2 Real scene	58
4.7	Discussion	62
4.8	Conclusion	63

5	Non-line-of-sight imaging	
	— Shape from radiance of three-bounce light paths	70
5.1	Related work	73
5.2	Problem statement	74
5.3	Analysis by synthesis optimization	74
	5.3.1 Differentiating transients	75
	5.3.2 Stochastic estimation and optimization	80
5.4	Geometry processing operations	83
5.5	Experiments	84
	5.5.1 Simulation results	85
	5.5.2 Real scene	86
5.6	Discussion and conclusion	87
6	Conclusions and Future Work	100
6.1	Future work	101
	Bibliography	103

List of Tables

2.1	Quantitative analysis on single-view reconstruction algorithm	15
3.1	Alternating image intensity threshold used for measuring two-bounce light paths	30
4.1	NLOS scene recovery related works	36
4.2	NLOS shape recovery results on planar objects	67
4.3	Recovery results with planes at different positions	67
4.4	Recovery results on curved objects	67
4.5	Recovery results on multi-object scene	69
6.1	Contributions	100

List of Figures

1.1	Real-world example of multi-bounce light path	2
1.2	Real-world objects with complex material properties	3
1.3	Three applications in this thesis	7
1.4	Thesis contributions	9
2.1	System overview for transparent object reconstruction	11
2.2	Geometric interpretation of Theorem 1	12
2.3	Geometric interpretation of Theorem 2	14
2.4	Illustration of Corollary 2	14
2.5	Synthetic experiment setup and results	17
2.6	Real experiment setup and results	18
3.1	System overview for shape from two-bounce light paths	21
3.2	Two-bounce light-path graph	22
3.3	Sample light-path graph	26
3.4	Synthetic experiment to compare single-bounce and two-bounce shape recovery	27
3.5	Visualization of scene shapes and materials	28
3.6	Example of rendered transients for light-path graph determination	29
3.7	Experiment results for shape from two-bounce light paths	29
3.8	Observable light paths	31
3.9	Light-path graph visualization	32
4.1	Non-line-of-sight imaging setup	34
4.2	Sample scene and rendered light transients	40
4.3	SNR of the first-returning photon compared to the full light transient	41
4.4	Stable estimate for first-returning photon	42
4.5	Statistical characteristics of the first-returning photon estimates	43
4.6	First-returning photon estimation is invariant to reflectance	44
4.7	Proof for Observation 1	45
4.8	Space carving visualization	46
4.9	First-returning photon is typically from a unique scene point	47
4.10	Uniqueness of shortest three-bounce light path	48
4.11	Visualization of Observation 2	48
4.12	Geometric interpretation of Observation 3	49

4.13	Normal estimation of the recovered NLOS scene points	50
4.14	Space carving for NLOS imaging	51
4.15	NLOS shape and normal under the local planarity assumption	52
4.16	Failure case for Algorithm 4	54
4.17	Quadratic models for NLOS imaging	55
4.18	Scene setup for space carving experiment	57
4.19	Space carving experiment result	58
4.20	NLOS shape recovery synthetic scene setup	59
4.21	NLOS shape reconstruction under varying scene reflectance	60
4.22	NLOS shape recovery with different noise levels	65
4.23	NLOS imagin setup	66
4.24	NLOS shape recovery with different integration time	66
4.25	NLOS shape estimation results with SPAD collected measurements	68
4.26	NLOS object tracking using a SPAD	69
5.1	NLOS surface optimization pipeline	73
5.2	Normal smoothness regularization	80
5.3	Geometry processing	81
5.4	Comparison of gradient estimation techniques	84
5.5	NLOS surface optimization results of multiple types of surface detail	91
5.6	NLOS surface optimization results with an alternative initialization method	92
5.7	Surface optimization using different numbers of measurements	93
5.8	NLOS surface optimization with simulated SPAD noise	94
5.9	Surface optimization progression	95
5.10	Recovering NLOS shape and reflectance	96
5.11	NLOS surface optimization using SPAD measurements	97

Chapter 1

Introduction

3D perception enables a wide spectrum of applications. From technologies of today, such as entertainment systems, and surveillance, to those in the near future, including augmented reality, autonomous driving, and 3D printing, shape estimation is inarguably a vital component of machine perception.

Classical techniques for shape estimation make several undesirable assumptions; in particular, direct path and Lambertian reflectance assumption. These assumptions are the keys to enable many classical computer vision techniques like multi-view geometry [47], shape from shading [53], structure from motion [125], photometric stereo [143], and structured light [113, 148]. However, when we examine the two assumptions in more detail, we can see that many common scenarios violate the assumptions. First, when a camera only observes direct light paths, i.e., light paths that only interact with the scene at a single location, different parts of the scene do not affect the measurement of each other and this greatly simplifies shape recovery. This assumption, however, omits many complex light-object interactions that are commonly seen in the real world, such as interreflections and subsurface scattering (see Figure 1.1). In the presence of such indirect light paths, each measurement made by the camera is linked to properties of multiple scene points. Shape estimation techniques need to disentangle the contributions from different parts of the scene, which is a hard problem. Second, when the objects in the scene exhibit Lambertian reflectance, their appearance does not change across views. This assumption is key to form correspondences across multiple camera views to estimate 3D geometry. However, real-world objects invariably violate the Lambertian reflectance assumption; for example, reflection on mirrors, caustics on a wine glass, the shine on metallic objects, and numerous phenomena (Figure 1.2). Without the assumption of Lambertian reflectance, shape estimation problem often becomes difficult since reflectance is also a variable in the image formation model. Shape estimation method that relies on image intensities will need to untangle the effect of geometry and reflectance.

In this thesis, we propose a new shape acquisition framework that particularly focuses on multi-bounce light paths. Instead of viewing multi-bounce light paths as undesirable effects that need to be eliminated during shape reconstruction, we turn them into a useful tool to understand scene properties. Each "bounce" in the light path is related to an instance of light-object interaction and it encodes rich information about the scene. The challenge is to find computationally tractable



Figure 1.1: **Real-world examples of multi-bounce light paths.** Light paths undergo at least two changes of direction when interacting with a transparent object.¹ Interreflections are prevalent when objects in the scene exist specular reflectance. Human skin is one common example of subsurface scattering.

techniques to utilize the information embedded in these multi-bounce light paths.

At the core of this thesis, we argue that light paths are better suited for shape estimation compared to images. Inspired from the work of Kutulakos and Steger [74], we use light paths as primitive to estimate an object's shape. Studying a single light path is akin to tracing the optical journey of a single photon. Each "bounce" is an optical event that can be described using simple physical laws; for example, mirrors alter the propagation direction of light at the surface, which follows the law of reflection. Transparent objects, on the other hand, change a light path's direction twice, each following Snell's law. Light transport is determined by the geometry and the BRDF of the scene. Each bounce of light path encodes properties of surface location, surface normal, and material properties at multiple scene points. By combining all different light paths, we gather rich information regarding the scene of interest.

There are many research efforts in dealing with multi-bounce light paths and objects with complex reflectance. We discuss these related works and show how some of them inspire us to use light paths as the building block for shape estimation

1.1 Related work

There are many results that focus on extending shape reconstruction techniques to reconstruct objects that exhibit interreflections or other complex material properties. We discuss two areas of research that are most related to this thesis. The first area focuses on isolating one-bounce light paths, also known as the direct global separation. The second deals with non-Lambertian reflectance, or even more complex scenario of spatially varying BRDF for shape reconstruction. We also briefly discuss time-of-flight imaging, which is one of the fundamental techniques that we use in this thesis.

¹Photo credit: Juinn-Kai Huang.

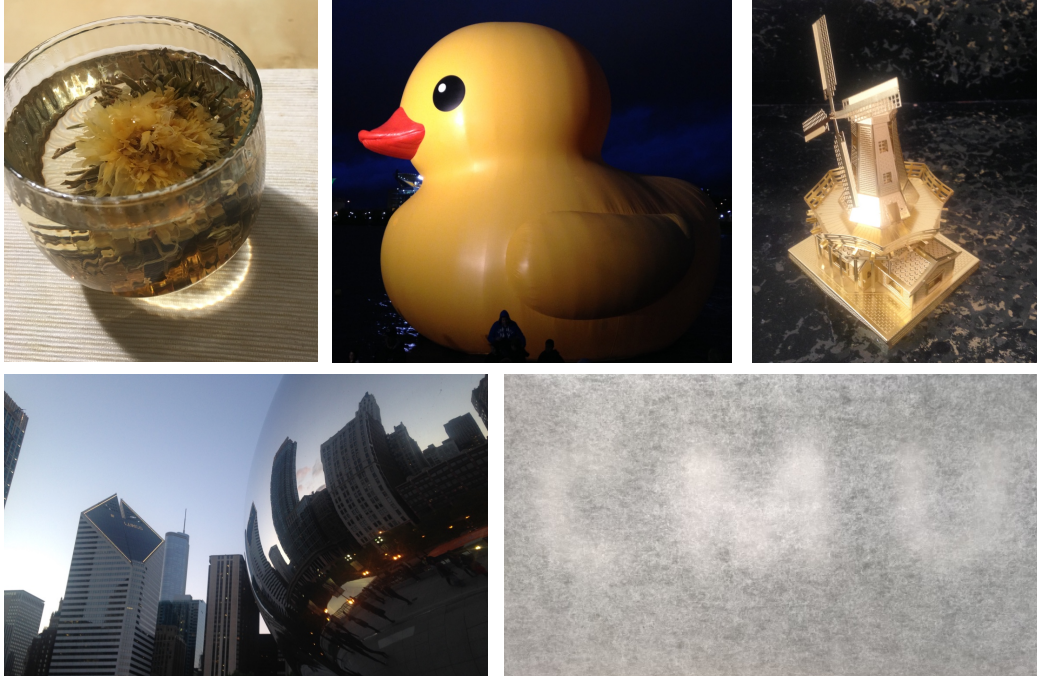


Figure 1.2: **Real-world objects with complex material properties.** Most objects in the real world have non-Lambertian reflectance. Glass, mirror, rubber, and metal all have drastically different appearances. Some materials create an opaque surface, for example, rubber, metal, and mirror. On the other hand, some materials are translucent, for example, a thin piece of paper and a glass of tea.

Direct global separation and global suppression. Most 3D reconstruction methods rely on the single-bounce light path assumption. Therefore, there are a lot of works that aim to extract the direct component of light transport or to suppress the effect of global component (multi-bounce light) so that the measurements are mainly from single-bounce light paths. Seitz *et al.* [115] show the existence of interreflection cancellation operator and show how it can decompose an image into the sum of n -bounce images. For a scene with Lambertian reflectance, the interreflection cancellation operator can be computed from the light transport matrix, thus direct light can be computationally recovered. Nayar *et al.* [91] propose a projector-camera system to illuminate the scene with high-frequency patterns to separate the direct components of light from global components.

Most methods in this field find methods and scenarios where the global effect is minimized. Gupta *et al.* [43] show that with an appropriate choice of spatial codes in structured light, the effects of subsurface scattering and interreflections can be minimized, resulting to a robust shape recovery for materials exhibiting moderate global light transport. O’Toole *et al.* [93] use the observation that one-bounce light paths need to follow epipolar geometry, and build a hardware prototype that only allows light travelling on epipolar plane. This imaging architecture decreases the impact of the global component of light transport.

For continuous wave time-of-flight, the global effect will affect the phase estimation, thus

global light needs to be taken care to give an accurate time-of-flight estimation, this is also called the multi-path interference problem. Gupta *et al.* [44] show that with the appropriate choice of temporal modulation frequencies in a ToF sensor, the global light transport effects can be significantly suppressed based on phasor analysis. Achar *et al.* [1] use the idea similar to [93] that direct component will follow epipolar geometry and build a time-of-flight imaging prototype.

While these methods focus on reducing the impact of multi-bounce light paths for shape estimation, in this thesis, we show that multi-bounce light paths actually contain rich information. Instead of treating multi-bounce light paths as noise, we explicitly study the geometric constraints of these light paths and use them for shape estimation.

Dealing with non-Lambertian reflectance. Without the assumption of Lambertian reflectance, finding the shape of an object through intensity measurements requires taking reflectance into account. Earlier attempts to deal with non-Lambertian reflectance aim to recover the shape of an object with homogeneous BRDF. A classical approach by Hertzmann and Seitz [52], called "Example-based photometric stereo", uses the observation that an object with a known shape made of similar material with the object of interest can provide information regarding how lighting changes the appearance of the object. By placing an object with similar material into the scene, the reflectance estimation problem can be avoided, thus solving the shape becomes simpler.

A special case of homogeneous BRDF is a mirrored object. We detail some works in this field since these relate to our work closely. A fundamental ambiguity in specular surface reconstruction is the so-called depth-normal ambiguity [63]. Given a light ray that is the reflection of a known 3D reference point on a mirror, the surface normal and the depth at which reflection occurs are constrained. Specifically, for every possible depth along the light ray, there is a corresponding normal leading to the same reference point. There are multiple approaches to resolve the depth-normal ambiguity for specular surfaces. Methods in [80, 122, 141] overcome the depth-normal ambiguity by using correspondences of a known reference target and by regularizing the depth using smoothness priors. In [10], the shape of a mirror is recovered by obtaining images of a planar target, thus the need for knowledge of the reference point is relaxed. Shape from specular flow [2, 112] recovers the shape of a mirror from scene motion, requiring little effort in terms of calibration. In this setting, it is even possible to obtain invariant signatures of the shape of the object from multiple images [111]. Reconstructing the shape of a mirror object is greatly simplified when thinking in terms of light paths instead of images. The problem for depth-normal ambiguity arises because of the additional two degrees of freedom of light ray direction from the reference points. Kutulakos and Steger [74] argue that light paths are more suitable for 3D reconstruction since depth estimation can be simplified to a light path triangulation problem. For a mirrored object, the surface of the object is exactly at the location of the "bounce" of the light path, which can be found by intersecting the incoming and outgoing light rays and the surface normal can be found by averaging the light ray directions. This simple yet powerful observation is central to many ideas in this thesis.

A step toward a more realistic setup is to deal with objects with spatially-varying BRDF. One approach to extend photometric stereo to deal with spatially-BRDF to solve a parametric BRDF model [35, 41] or assume the BRDF to be a linear combination of exemplar BRDFs [3, 54]. Another approach is to find reflectance-invariant features by considering object and camera motions

[16, 17, 18, 140]. Since photometric stereo methods rely on measured intensities, jointly solving shape and reflectance or finding reflectance-invariant features are required to deal with objects with complex materials.

The aforementioned methods only deal with opaque objects. Another type of complex reflectance is translucent materials, for example, a glass of milk, soap, human skin, etc. Scattering effect needs to be taken into account and the light path of interest is typically a multi-bounce light path. To help understanding translucent object properties, [39, 40, 89] measure material properties of objects with a known shape. Even though the object shape is known, estimating material property is non-trivial and requires complex imaging setups or computational tools. To simultaneously recover material and shape requires simplification of the image formation model, for example, assuming a one-bounce scattering light path [55], or using tomography for shape recovery [42, 76, 116].

More related to our work is a simpler instance of translucent objects, namely, transparent objects. For transparent objects, the light path of interest is invariably multi-bounce but is typically limited to two bounces. The main challenge of reconstructing a transparent object is then to recover the two refractive events that happen at two distinct unknown locations. Most methods in this field utilize that Snell’s law is the underlying physics that determine the appearance of a transparent object. For thin objects, we can even approximate the light path with a single-bounce light path [78, 142, 145] and recover its shape. Kutulakos and Steger [74] propose a theoretical analysis on utilizing the geometric properties of light paths and show that if we have location and direction of a light path before and after interacting with the transparent object, it is possible to reconstruct its shape. This analysis of turning physical phenomena happening at each bounce of the light path to geometric constraints of the object becomes the foundation of this thesis. This thesis can be viewed as an extension of the analysis in [74] to a broader class of objects and with ToF information of the light paths.

Time-of-flight imaging. Our proposed shape estimation framework builds on the fact that we can measure different properties of light paths, such as position, direction, radiance, and time of flight. Position, direction, and radiance can be acquired by typical cameras or light field cameras. Time-of-flight imaging is a more recent technology that uses transient radiometric sensors [59] to gather time-resolved intensity measurements. Since the speed of light is large (3×10^8 m/s), the associated time scales of travel times are invariably very small, typically picoseconds for the scale of everyday objects. The time of flight provides the total path length the photon takes from the light source to the scene and potentially bounces at multiple locations then back to the sensor. That is, if we send a short pulse of light from the light source, we want to recover the radiance of each light path returning to the sensor at different instants of time.

One of the most commonly seen time-of-flight technologies is LIDAR, which is widely used for autonomous driving. LIDAR sends a short pulsed laser beam into the scene and measures the time-of-flight of the returning light. The time of flight and the direction of the beam provide the estimation of a single 3D point of the scene. By changing the direction of the laser beam, LIDAR provides a dense 3D point cloud as scene representation. More recently, single photon avalanche diodes (SPADs), which is the underlying technology of LIDAR, provides an alternative approach to measure the time of flight of light paths. The SPAD is operated under infinite gain and saturates

whenever a photon incidents on the sensor; the time-stamp of this event measures the ToF of the photon. Buttafava *et al.* [15] show that, when illuminating the scene with an impulse train, the histogram of photon arrivals is a good approximation to the light transient. O’Toole *et al.* [94] further show an architecture using a line SPAD to acquire the transient image of the whole scene more efficiently. Gariepy *et al.* [33] use a SPAD array to acquire the light in flight in a 2D field of view. Sun *et al.* [121] further improve spatial resolution by using compressive sensing techniques. To account for SPAD characteristics while imaging, Hernandez *et al.* [51] provide analysis and simulation results that link the true light transients to the SPAD readouts. Unlike LIDAR that provides only a single depth estimation of the light path, a SPAD sensor can be used to estimate the light transient response. The time-resolved radiometric information provides new opportunities in hard scenarios, such as non-line-of-sight imaging and scattering. Another popular hardware setup to measure the light transient is a streak camera. A streak camera uses a time-varying electric field to introduce a time-dependent spatial displacement to incoming photons. The time-of-flight information of one scan line of the scene is acquired in one capture session. Velten *et al.* [136] use this technology to image at time resolutions of tens of picoseconds.

An alternate approach to measure the light transient is to illuminate the scene with an amplitude modulated source and measure the amplitude and phase shifts in the amplitude wave caused by light propagation using a photonic mixer device (PMD) [48]. The PMD measurements can be interpreted as the Fourier coefficients of the transient response at the frequency of the illumination; hence, by illuminating the scene with different frequencies, the light transient can be acquired by measuring and inverting the Fourier coefficients [104]. This approach, however, only provides temporal resolution in nanoseconds due to the limited modulation frequency of the LED light source (typically 10-100MHz). Also limited to the number of sampled frequency, the light transient appeared to be smoothed. By including more light sources and PMDs, Shrestha *et al.* [117] show several applications that measure the interference of multiple light sources. Interference of multiple light paths can be used to further increase the temporal resolution. [37, 60] utilize interference of coherent wave to find the path length of light paths. In [37], they build a prototype that uses the Michelson interferometer setup to find the depth of the object of interest. By using a translation stage to control the reference light path length, the path length of the light path that interacts with the object can be found by the interfering with the reference light path. However, the acquisition time is very large since the resolution is controlled by the number of sampled depth using the translation stage. More recently, Dynamic Heterodyne Interferometry [84] and Superheterodyne Interferometry [77] bridge the resolution gap between continuous-wave ToF (MHz) and interferometry-based system (THz), achieving resolution that is GHz. The key is to have two coherent light sources having slightly different wavelengths to superimpose with each other. The superimposed light will create a synthetic frequency, roughly GHz, and can be measured with the optical beat note measured at the sensor. Maeda *et al.* [84] further polarize the light source so that four measurements can be simultaneously measured to estimate the phase shifts.



Figure 1.3: **Three applications in this thesis.** We show shape reconstruction for transparent objects, concave objects, and non-line-of-sight objects.

1.2 Contributions

The key hypothesis of this thesis is that shape estimation is significantly simplified if we study attributes of light paths and explicitly utilize their properties, including position, direction, time of flight, and radiance. In particular, by considering properties associated with a single light path, we can link physical phenomenon on each instance of light-object interaction. We then can use these physical phenomena to find geometric and physically accurate techniques for shape recovery to deal with multi-bounce light paths interacting with objects with complex reflectance. The techniques in this dissertation are most effective when dealing with some of the most challenging problems in 3D estimations. We demonstrate the effectiveness of the hypothesis by three different applications: transparent shape reconstruction, concave shape reconstruction, and non-line-of-sight (NLOS) imaging (Figure 1.3). In these three cases, the light paths undergo at least two bounces and the object for shape recovery can be non-Lambertian.

- **Transparent object reconstruction with light ray correspondences.** Glass is a material that we commonly use in numerous applications in our daily lives. However, traditional 3D reconstruction methods cannot handle glass objects well, since its appearance depends heavily on the surrounding environment and the shape of the object. It is also difficult to use a laser to scan a transparent object, since only a fraction of light will directly reflect back to the sensor. Most light will undergo at least two changes of direction; once entering and once exiting the transparent object. A theoretical analysis made by Kutulakos and Steger [74] provides a systematic way to study light interacting with mirrors or glasses. By explicitly exploiting the law of reflection and the law of refraction, we can use the direction changes of light paths to form geometric constraints of the object of interest. We re-parameterize the proof shown in [74] and show two ambiguities pertaining surface depth and normal for transparent objects. This re-parameterization enables enforcing surface priors, such a locally planar, more directly. We show a single-view shape recovery algorithm that assumes local planarity on one side of the transparent object.
- **Concave object reconstruction with ToF of two-bounce light paths.** For a concave object, since several facets of the scene are directly visible to each other, interreflection happens fre-

quently. For any shape-from-intensity method [53, 143], if interreflection is not taken into account in the image formation model, the estimation will be erroneous. An early work by Nayar *et al.* [90] incorporated interreflection of a Lambertian surface into the forward model and proposed an iterative solver to optimize the shape. However, this technique requires the assumption of a Lambertian surface so that the image formation model can be simplified. In this work, we explore an alternative method to deal with interreflections that do not rely on the Lambertian assumption. We measure the ToF of two-bounce light paths, which provides many beneficial properties; for instance, ToF is a quantity that relates to the geometry of the object. As long as the light path is measurable, the ToF value is not affected by the material reflectance. We can then use the ToF of two-bounce light paths to form geometric constraints. These geometric constraints will relate the depth of two distinct scene points to a measured quantity. We show a systematic analysis of these constraints to establish a unique shape recovery guarantee.

- **Non-line-of-sight imaging with ToF and radiance of three-bounce light paths.** The capacity of multi-bounce light paths really shines when looking at a recently popular field, non-line-of-sight imaging. NLOS imaging aims to recover properties of an object that is not in the observer’s field of view using measurements gathered from within the field of view. The main challenge is that the direct component of light transport contains no information regarding the hidden object and only three- or higher-order- bounce light paths can potentially interact with the object of interest. Typically, the measurement of three-bounce light paths is contributed by multiple light paths of the same length. Finding the shape of the hidden object requires untangling contributions from different scene parts. We develop two simple and physically accurate methods to deal with this complex imaging problem. First, we observe that a subset of the measurements, which we called the first-returning photons, contains geometric properties that are useful for shape recovery. We derive a shape estimation method that is based on Fermat’s principle to recover the 3D points and their normals on the hidden object. Second, we revisit the image formation model for three-bounce light paths. Unlike previous methods in this field that modify the image formation model for computational tractability, we develop a surface optimization framework that works with the physically accurate image formation model. We develop an inverse rendering pipeline that relies on Monte Carlo rendering for both the forward model and its gradient with respect to the surface.

We tackle some of the hardest problems in 3D reconstruction using computationally tractable and physically accurate methods (see Figure 1.4). While we show three applications that involve multi-bounce light paths under the visible lights bandwidth, there are a lot of similarities in all types of wave propagation. The ideas presented in this dissertation are applicable to other disciplines, for example, acoustics [25], medical imaging (Magnetic Resonance Imaging, Computed Tomography), and wifi localization, since they share similar underlying imaging principles as the three applications mentioned in this dissertation.

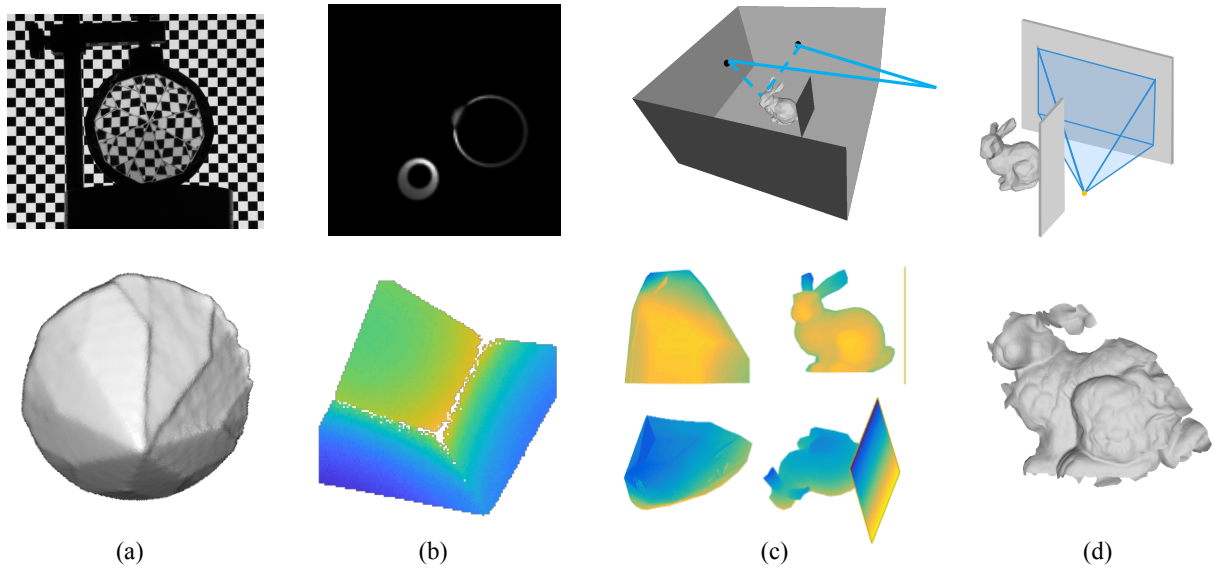


Figure 1.4: **Thesis contributions.** We tackle some of the most complex shape estimation problems that involve multi-bounce light paths. (a) Transparent objects change the propagation direction of light at least two times. Estimating the shape of transparent objects requires simultaneously recover the geometric properties at two locations. We utilize Snell’s law to create geometric constraints of the light paths and subsequently, the location of the bounce can be found by incorporating surface local planarity prior. (b) Interreflection is a common phenomenon in concave objects. The measurements contain information about multiple scene points. By utilizing time-of-flight imaging, light paths of different path lengths can be separated. The path length of a two-bounce light path creates a geometric constraint of possible depth values. With multiple path-length information regarding two-bounce light paths, we show a scene depth estimation method that makes inter-reflection no longer an undesirable effect. (c) When an object is occluded from view, the only light path that interacts with the object is at least three bounces. With time-of-flight imaging, we utilize the geometric constraints of the shortest light path to find the spatial extent of the hidden object. Note that shortest light paths will follow Fermat’s principle, thus additional geometric constraints can be formed for scene point localization. (d) Similar to (c), we study another setup of non-line-of-sight imaging. We show the physically accurate rendering equation for time-resolved radiometric measurements. We find the expression of the gradient of the rendering equation with respect to the hidden surface and devise an inverse rendering pipeline. We use stochastic gradient descent method to optimize the hidden surface so that the rendered time-resolved radiometric measurements become consistent with the measurements.

Chapter 2

Transparent object reconstruction — Shape from deflection of two-bounce light paths

One of the most challenging objects for 3D reconstruction is transparent and specular objects, which do not have their own appearance. We can only observe how the objects distort the surrounding environment, that is, we can only capture how a light ray deflects after interacting with the object of interest. A seminal result [74] suggests the use of correspondences of light rays to infer the geometry of the object. They show that the object can be reconstructed, provided that the direction of a light path does not change more than two times.

Establishing light ray correspondences is a direct result of tracing a light path. We can illuminate the scene with a single light ray, and then capture the light ray after interacting the scene. This way, we capture how the object changes the light path’s propagation direction. In practice, this process can be achieved by a camera-projector system such as the setup used in [74] or a light-field probe [142] that encodes directional information in the color of the light ray.

While considering light ray in space, every optical event with a mirror or a transparent object will follow the physical laws of reflection and refraction. There are three keys to shape recovery for transparent objects:

- **Light paths are connected.** If a scene point is observable by the camera, there exists at least one light path that originates from the light source and ends up in the camera.
- **Helmholtz reciprocity principle.** The reverse of the light path encounters the same optical events as of the forward light path.
- **Fermat’s principle.** The path between two points taken by light is the one with the smallest propagation time. In particular, Fermat’s principle explains the law of reflection and Snell’s law.

We show that these three simple physical laws can lead to geometric constraints of light paths and subsequently lead to a shape estimation algorithm [129]. We address the following problem in refractive shape estimation: given a single light-ray correspondence, what shape information of the transparent object is revealed along the path of the light ray, assuming that the light ray refracts twice (see Figure 2.1). We characterize the shape estimation by studying the ambiguities in the

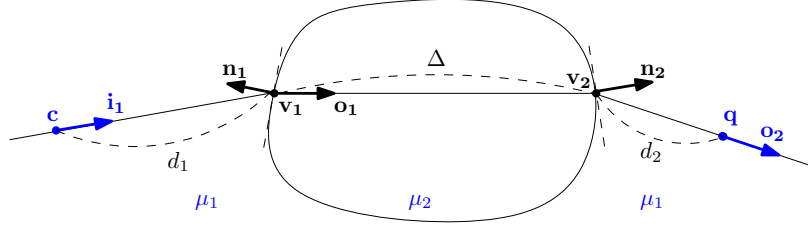


Figure 2.1: **Shape estimation of a transparent object.** The light path of interest originates from \mathbf{c} with orientation \mathbf{i}_1 , after interacting with the object (refracting twice), ends up being a light ray crossing \mathbf{q} with orientation \mathbf{o}_2 . From this information of the two-bounce light path, we seek to understand the shape constraints of the transparent object. The variables labeled in blue are known and form the inputs to the problem.

depth and surface normal along a single light path.

2.1 Problem statement

Notation. We denote vectors in **bold**. Light rays in 3D are denoted by a pair of vectors, $\{\mathbf{o}, \mathbf{p}\}$, where $\mathbf{o} \in \mathbb{R}^3$ is a unit-norm vector that denotes the orientation of the ray and $\mathbf{p} \in \mathbb{R}^3$ is a point in 3D space that the ray passes through. For $\lambda \in \mathbb{R}$, $\{\mathbf{o}, \mathbf{p}\}$ and $\{\mathbf{o}, \mathbf{p} + \lambda \mathbf{o}\}$ are identical.

Suppose that a ray $\{\mathbf{i}_1, \mathbf{c}\}$ is incident on a transparent object and after refracting twice, once each upon entering and exiting the object, becomes the ray $\{\mathbf{o}_2, \mathbf{q}\}$ (see Figure 2.1). Given knowledge of the rays $\{\mathbf{i}_1, \mathbf{c}\}$ and $\{\mathbf{o}_2, \mathbf{q}\}$, the relative refractive indices of the medium μ_1 and the object μ_2 and their ratio $\rho = \mu_2/\mu_1$, what can we infer about the shape of the object as encoded in the locations of the refraction events, \mathbf{v}_1 and \mathbf{v}_2 , and the surface normals, \mathbf{n}_1 and \mathbf{n}_2 ?

2.2 Shape from light-ray correspondences

We answer this question by first finding the physical laws associated with the light path being refracted twice. We conclude the geometric constraints in terms of the following three observations:

Observation 1. Since the refraction events occur on the light rays, we can identify two depth values d_1 and d_2 such that $\mathbf{v}_1 = \mathbf{c} + d_1 \mathbf{i}_1$ and $\mathbf{v}_2 = \mathbf{q} - d_2 \mathbf{o}_2$. The problem is equivalent to estimating d_1, d_2, \mathbf{n}_1 , and \mathbf{n}_2 .

Observation 2. The light path is fully determined from the depth and surface normals at \mathbf{v}_1 (or equivalently, \mathbf{v}_2). Given d_1 and \mathbf{n}_1 , the outgoing ray \mathbf{o}_1 is fully specified from the laws of refraction. The intersection of this ray with $\{\mathbf{o}_2, \mathbf{q}\}$ provides both the surface normal \mathbf{n}_2 and the 3D point \mathbf{v}_2 .

Observation 3. Any constraint that we derive on the shape at \mathbf{v}_1 , the first refraction point, translates to a similar constraint on the shape at \mathbf{v}_2 . This is simply a consequence of Helmholtz reciprocity.

Our main results are in the form of ambiguities in the values of the depth d_1 given knowledge of the surface normal \mathbf{n}_1 , and vice versa. These can be succinctly summarized in the following statements.

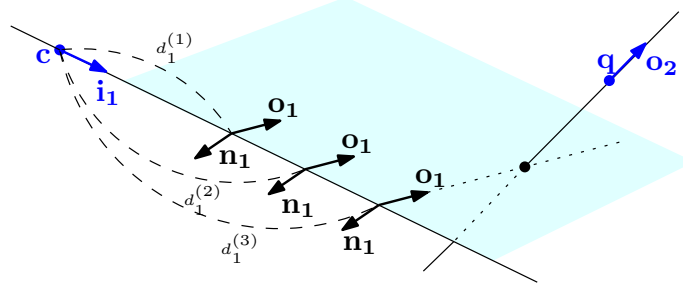


Figure 2.2: **Illustration of Theorem 1.** Given normal \mathbf{n}_1 , \mathbf{o}_1 is fully determined. By finding the intersection of the light ray $\{\mathbf{o}_2, \mathbf{q}\}$ and the plane formed by all possible \mathbf{v}_1 and \mathbf{o}_1 , we can find the corresponding depth $d_1^{(3)}$.

Theorem 1 (Depth ambiguity given normal). *Given the surface normal \mathbf{n}_1 , the depth d_1 at which the refraction occurs is unique, provided the light path does not entirely lie in a plane.*

Proof. By Snell's law, the refracted ray \mathbf{o}_1 is uniquely determined given both the incident ray \mathbf{i}_1 and the normal \mathbf{n}_1 by

$$\mathbf{o}_1 = \frac{\mathbf{i}_1}{\rho} - \left(\langle \mathbf{i}_1, \mathbf{n}_1 \rangle + \sqrt{\left(\langle \mathbf{i}_1, \mathbf{n}_1 \rangle^2 - (1 - \rho^2) \right)} \right) \frac{\mathbf{n}_1}{\rho}. \quad (2.1)$$

Since a light path is connected, the relationship between d_1 and d_2 can be characterized as

$$\mathbf{c} + d_1 \mathbf{i}_1 + \Delta \mathbf{o}_1 + d_2 \mathbf{o}_2 = \mathbf{q}.$$

This can be further simplified to

$$d_1 \langle \mathbf{i}_1 \times \mathbf{o}_2, \mathbf{o}_1 \rangle = \langle (\mathbf{q} - \mathbf{c}) \times \mathbf{o}_2, \mathbf{o}_1 \rangle.$$

When $\langle \mathbf{i}_1 \times \mathbf{o}_2, \mathbf{o}_1 \rangle \neq 0$,

$$d_1 = \frac{\langle (\mathbf{q} - \mathbf{c}) \times \mathbf{o}_2, \mathbf{o}_1 \rangle}{\langle \mathbf{i}_1 \times \mathbf{o}_2, \mathbf{o}_1 \rangle}. \quad (2.2)$$

In (2.2), \mathbf{i}_1 , \mathbf{o}_2 , \mathbf{c} and \mathbf{q} are known. Further, the value of \mathbf{o}_1 is fully-determined from (2.1) when \mathbf{n}_1 is given. Therefore, provided $\langle \mathbf{i}_1 \times \mathbf{o}_2, \mathbf{o}_1 \rangle \neq 0$, d_1 is uniquely determined when the surface normal \mathbf{n}_1 is known. Finally, $\langle \mathbf{i}_1 \times \mathbf{o}_2, \mathbf{o}_1 \rangle = 0$ if and only if \mathbf{o}_1 lies in the plane spanned by \mathbf{i}_1 and \mathbf{o}_2 or, equivalently, the entire light path lies in a plane. \square

We provide a geometric interpretation of Theorem 1 in Figure 2.2. Given \mathbf{i}_1 and \mathbf{n}_1 , \mathbf{o}_1 is determined. All possible rays after the first refraction will form a plane. The light ray $\{\mathbf{o}_2, \mathbf{q}\}$ will intersect the plane at one point. Therefore, we can find a unique depth d_1 .

Corollary 1 (Planar light path). *If the vectors \mathbf{i}_1 , \mathbf{o}_2 and $(\mathbf{q} - \mathbf{c})$ are co-planar, then specifying \mathbf{n}_1 does not constrain d_1 .*

Theorem 2 (Normal ambiguity given depth). *Given the depth d_1 , the surface normal \mathbf{n}_1 is constrained to lie on a 1D curve, which is the intersection of an oblique cone with a unit sphere.*

Proof. Recall that the first refraction happens at \mathbf{v}_1 , which is d_1 units away from the camera center,

$$\mathbf{v}_1 = \mathbf{c} + d_1 \mathbf{i}_1.$$

The second refraction occurs at \mathbf{v}_2 which lies on the line defined by a point \mathbf{q} and the vector \mathbf{o}_2 ,

$$\mathbf{v}_2 \in \{\mathbf{q} - \lambda \mathbf{o}_2, \quad \lambda \in \mathbb{R}, \lambda \geq 0\}.$$

From Figure 2.3, we observe that \mathbf{o}_1 belongs to the plane that $\mathbf{q} - \mathbf{v}_1$ and \mathbf{o}_2 are on. Therefore, we can represent \mathbf{o}_1 by:

$$\mathbf{o}_1 = \mathcal{B}(d_1) \begin{bmatrix} \cos(\psi) \\ \sin(\psi) \end{bmatrix}, \quad \psi \in [0, 2\pi),$$

where $\mathcal{B}(d_1)$ is an orthonormal basis for the column span of $\mathbf{q} - \mathbf{v}_1$ and \mathbf{o}_2 .¹ Since \mathbf{c} , \mathbf{i}_1 , \mathbf{o}_2 and \mathbf{q} are known and $\mathbf{v}_1 = \mathbf{c} + d_1 \mathbf{i}_1$, \mathcal{B} is dependent only on d_1 .

We define \mathbf{n}_\perp as a unit-norm vector, co-planar to \mathbf{i}_1 and \mathbf{n}_1 , and orthogonal to \mathbf{n}_1 . From Snell's law, we know $\mu_1 \sin \theta_1 = \mu_2 \sin \theta_2$, where θ_1 and θ_2 are the angles formed by \mathbf{i}_1 and \mathbf{o}_1 , respectively, to \mathbf{n}_1 .

$$\begin{aligned} \mu_1 \langle \mathbf{i}_1, \mathbf{n}_\perp \rangle &= \mu_2 \langle \mathbf{o}_1, \mathbf{n}_\perp \rangle \\ \mu_1 \left\langle \mathbf{i}_1 - \frac{\mu_2}{\mu_1} \mathbf{o}_1, \mathbf{n}_\perp \right\rangle &= 0 \\ \langle \mathbf{i}_1 - \rho \mathbf{o}_1, \mathbf{n}_\perp \rangle &= 0 \end{aligned}$$

Recall that \mathbf{n}_1 is co-planar to \mathbf{i}_1 and \mathbf{o}_1 and, by definition, $\langle \mathbf{n}_1, \mathbf{n}_\perp \rangle = 0$. Hence, we can conclude that \mathbf{n}_1 is parallel to $\mathbf{i}_1 - \rho \mathbf{o}_1$.

$$\mathbf{n}_1 \propto \mathbf{i}_1 - \rho \mathcal{B}(d_1) \begin{bmatrix} \cos(\psi) \\ \sin(\psi) \end{bmatrix}, \quad \psi \in [0, 2\pi). \quad (2.3)$$

The RHS of (2.3) traces a circle in \mathbb{R}^3 as ψ varies. Since $\|\mathbf{n}_1\| = 1$, we can recover \mathbf{n}_1 by computing the intersection of an oblique cone and a unit sphere. Therefore, normal \mathbf{n}_1 lies on a closed 1D curve on a 2D unit sphere. \square

Corollary 2 (Proposition 1a of [74]). *To uniquely identify the surface normal at a point, we need at least three light-ray correspondences.*

Proof. Given the point under consideration, we know the value of d_1 for each ray correspondence. From Theorem 2, each light-ray correspondence restricts the surface normal to lie on a closed 1D curve. However, any two arbitrary 1D curves on the unit-sphere can potentially intersect. Hence, we need a third correspondence to verify the intersection produced by the first two correspondences (see Figure 2.4). \square

¹Another way to parameterize is $\mathbf{o}_1 = \frac{\mathbf{q} - \lambda \mathbf{o}_2 - \mathbf{v}_1}{\|\mathbf{q} - \lambda \mathbf{o}_2 - \mathbf{v}_1\|}$, where $\lambda > 0$. While this constraints the surface normal to a smaller set, the resulting expressions are harder to analysis due to their complex dependence on λ .

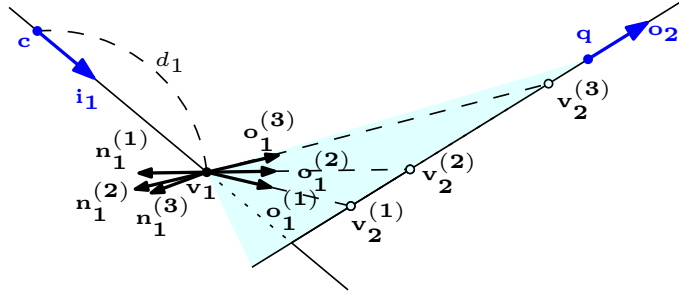


Figure 2.3: **Illustration of Theorem 2.** The refracted ray \mathbf{o}_1 lies in the plane spanned by $\mathbf{q} - \mathbf{v}_1$ and \mathbf{o}_2 . By choosing different \mathbf{v}_2 locations, we will have different corresponding \mathbf{o}_1 and \mathbf{n}_1 . In this figure, we show the normal $\mathbf{n}_1^{(i)}$ and $\mathbf{o}_1^{(i)}$ corresponding to different $\mathbf{v}_2^{(i)}$ locations.



Figure 2.4: **Illustration of Corollary 2.** (a) Any two 1D curves on the sphere can intersect. (b) We need a third curve to validate if the intersect is indeed correct.

Remark. The depth-normal ambiguities described in Theorems 1 and 2 are fundamental to studying the shape of transparent objects, where we can expect a majority of light-ray correspondences to be from double-refraction events. These are similar in spirit to two well-known ambiguities in computer vision: the depth-normal ambiguity for mirror object, and the bas-relief ambiguity in Lambertian shape-from-shading. An understanding of these fundamental ambiguities is important to the design of techniques for shape estimation.

Relationship to [74]. In Kutulakos and Steger [74], the algorithmic development, as well as analysis, is performed using d_1 and d_2 — variables pertaining to two distinct points, \mathbf{v}_1 and \mathbf{v}_2 . In contrast, we only use d_1 and \mathbf{n}_1 , which are local to \mathbf{v}_1 . This leads to a simpler explanation of the underlying ambiguities that we state in Theorems 1 and 2.

Algorithm 1 (Single-view reconstruction algorithm using surface prior).

If the first refraction happens on a plane, \mathbf{n}_1 is the same for all the light-path correspondences. Thus, from Theorem 1, for every candidate \mathbf{n}_1 , we can calculate d_1 of m light-ray correspondences. Since we impose a planar model on the collection of \mathbf{v}_1 , we use the variance of

$$[\langle \mathbf{n}_1, \mathbf{v}_1^{(1)} \rangle, \langle \mathbf{n}_1, \mathbf{v}_1^{(2)} \rangle, \dots, \langle \mathbf{n}_1, \mathbf{v}_1^{(m)} \rangle]$$

to determine the goodness the candidate \mathbf{n}_1 . The normal that gives the smallest variance is the normal estimation. Once the normal is recovered, both \mathbf{v}_1 and \mathbf{v}_2 are determined.

View number	1	2	3	4	5	6	7
Elevation (°)	84.0	84.0	82.6	79.4	83.5	84.2	83.5
Azimuth (°)	82.7	82.9	92.5	68.9	63.2	69.8	61.4
Standard deviation (°)	3.9	0.7	1.9	1.9	3.8	0.9	6.3

Table 2.1: **Normal estimation using different camera views.** We randomly select 50 light-ray correspondences to estimate each normal. Shown are average normal estimates in terms of azimuth and elevation angles as well as standard deviation across 50 trials.

2.3 Experiments

Simulated results. The acquisition for the ground truth of depth and normal of transparent objects is difficult. We use synthetic data and add different levels of noise to evaluate the effectiveness and robustness of our proposed single-view algorithm.

In our experiment, we set the first surface as a plane and the second surface some arbitrary shapes (see Figure 2.5). The refractive index is set to 1.2. We only collect light-ray correspondences that have double refraction and neglect light rays corresponding to total internal reflection. As shown in Figure 2.5 (b)-(e), by enforcing the planar model on one side, we can recover a complex object with low error.

To evaluate the robustness of the algorithm, we perturb \mathbf{i}_1 and \mathbf{o}_2 with different SNR values. We repeat the experiment five times and average the reconstruction error. Results in Figure 2.5 show that even with the presence of noise, our method still gives a small error.

Real scene. We use the real-world dataset from [74] to verify our single-view reconstruction algorithms. The data is collected by placing the object of interest between a camera and a movable LCD, as shown in Figure 2.6 (a). Image pixel to LCD pixel correspondences are collected at two LCD positions to give light-ray correspondences.

We observe that the back of the diamond scene is planar, therefore, we impose a planar model on the back of the diamond. The refractive index is set to 1.55. As shown in Figure 2.6 (b), by enforcing the planar model on one side, we can recover a complex object. Without enforcing the additional constraint, different facets of the diamond meet at the same position in space. To evaluate the correctness of the normal estimation, we use different camera views. The normal estimation should be consistent in all views. The results are shown in Table 2.1.

2.4 Conclusion

We demonstrate how to use light-ray correspondences of transparent objects for shape recovery. By capturing how a light ray changes its direction in space, the refraction events are linked to geometric constraints of the shape. From Snell’s law, two novel depth-normal ambiguities for transparent objects are found, which is similar in the spirit of Bas relief [7] in Lambertian objects and depth normal ambiguity [63] in mirrored objects. We further studied how simple surface

priors, such as a planar model, can be incorporated into a single-view shape estimation algorithm for transparent objects.

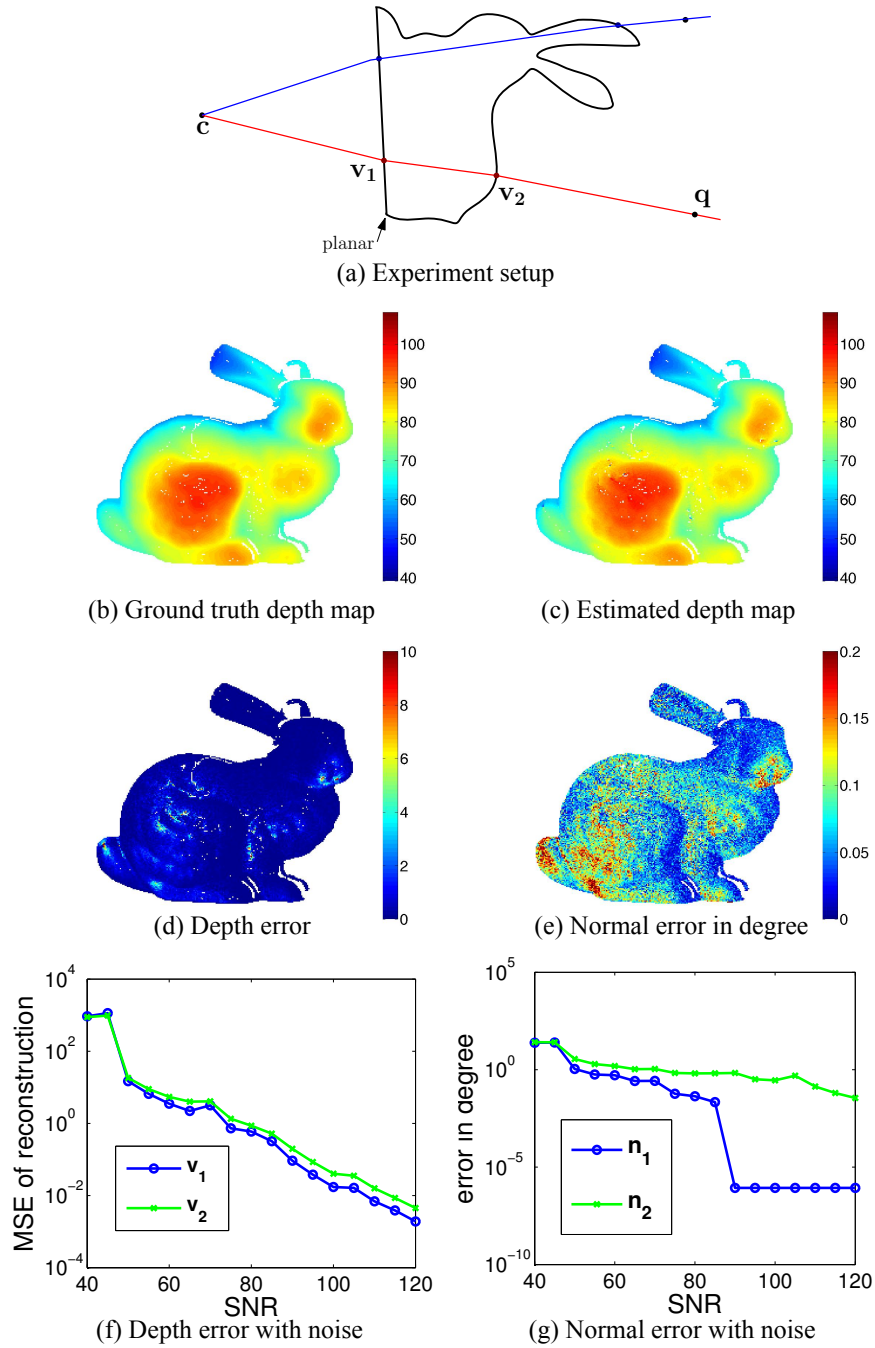
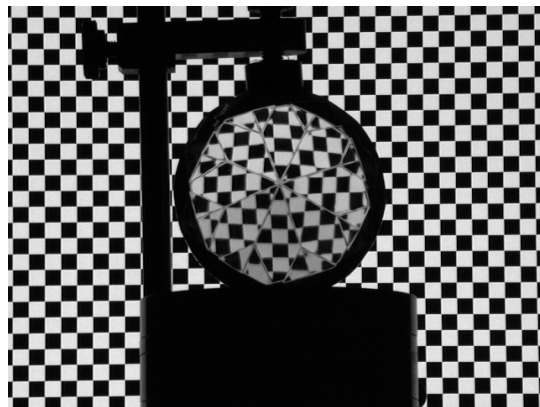
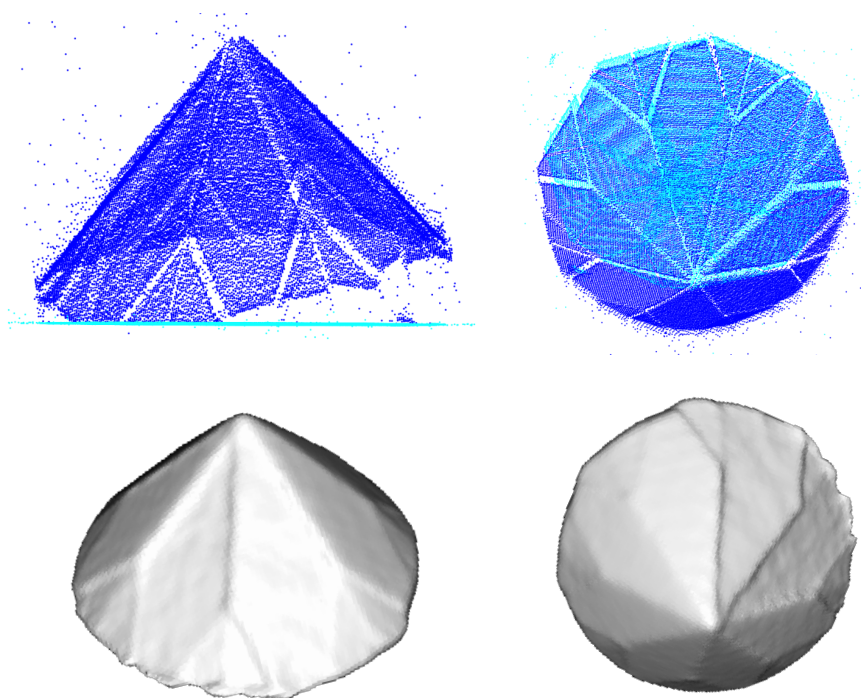


Figure 2.5: **Synthetic experiment setup and results.** (a) The object that we used in this experiment is planar on one side and a complex surface (bunny) on the other side. In (b)-(e), we show the performance of our method on the complex side. The light-ray correspondences are generated with ground truth then perturbed by SNR=80. In (f) and (g), we show the error with respect to different SNR values.



(a) Experiment setup



(b) Estimated point cloud and surface

Figure 2.6: **Real experiment setup and results.** (a) The dataset is courtesy of Kutulakos and Steger [74]. (b) We exploited the planarity of the “back” of the diamond to recover its shape from light rays observed at a single camera. Specifically, we randomly chose 40 light rays (out of a total of 33,737) entering the camera to estimate the normal and intercept of the plane. Given the parameters of the plane, we can now estimate densely both the front surface (shown in blue) and the back (shown in cyan). We also show the front side mesh visualization.

Chapter 3

Concave object reconstruction — Shape from time-of-flight of two-bounce light paths

Mirrored and transparent objects are two classes of non-Lambertian objects that have concise physical models on how they interact with light rays. By using light-ray correspondences and the associated physical laws, the shape of the object can be recovered. However, we cannot directly apply Fermat’s principle to find geometric constraints of objects with arbitrary reflectance, since the light transport of non-Lambertian objects is a function of both the shape and the reflectance of the composition materials. Therefore, we need to estimate shape and reflectance function at the same time, which is extremely challenging, in particular, due to two properties: 1) reflectance affects the image formation; 2) both shape and reflectance functions are high-dimensional.

We solve this problem by incorporating additional information about light paths – the time taken by a photon to traverse the light path. The time of flight of a photon is valuable information since the path length is a function of the geometry of the scene; we can avoid the need for jointly estimating geometry and reflectance functions. That is, a technique that requires only the path length information can work for scenes with arbitrary reflectance functions including mirrors, spatially-varying BRDF and transparent objects.

We show how to recover shape from light paths that undergo two changes of direction [130]. Unlike the approaches of existing active depth sensing techniques (LIDAR, ToF, radar) that rely on single-bounce light paths, we harness information of shape from two-bounce light paths. In this chapter, we show that two-bounce light paths can be sufficient for shape reconstruction. Also, there exist many advantages of two-bounce light paths:

- *The number of two-bounce light paths.* If there are N scene patches, there are at most N direct light paths, but the scene can potentially create $\binom{N}{2}$ two-bounce light paths. By considering two-bounce light paths, we create an over-determined system that is more robust to invert.
- *Generalization to arbitrary material and shape.* Specular and shiny materials (metals, liquids, glass) – in general, reflective surfaces – concentrate light along certain directions. This often results in the absence of single-bounce light paths. Here, the use of two and higher-order bounces

can alleviate the lack of information in the single-bounce light paths. Also, for a concave object, most of the light paths will interact with the object at multiple locations, therefore, it is hard to observe a single-bounce light path. Most of the observable light paths will be beyond two bounces.

- *Alternative imaging setup.* The results shown in this chapter indicate that we can uniquely recover shape from two-bounce light paths. This brings a new imaging strategy. We can image the scene multiple times with only one point of the scene being illuminated. We repeat this process by redirecting the illumination beam to another scene point until the sufficient condition of unique surface recovery is satisfied. Finally, we can use the algorithm developed in this chapter to recover the shape of the whole scene.

We establish theoretical and algorithmic foundations for shape estimation from two-bounce light paths [130], i.e., light paths where photons from a light source interact with the scene exactly twice before reaching the sensor.

3.1 Related work

There has been very little prior work in understanding the relationship between scene shape and the associated multi-bounce light paths. Most related to ours are two works that study the inter-reflections. An earlier approach by Nayar *et al.* [90] notice the presence of interreflection degrades the results from photometric-based methods. They estimate the shape of a concave object iteratively while progressively accounting for interreflections. Liu *et al.* [81] recover the shape of a Lambertian scene from the form factor of two-bounce light paths. By using local planarity and common elements in different two-bounce light paths, the scene geometry can be recovered without scaling ambiguity. A major problem in the two works is that the scene reflectance is assumed to be Lambertian so that the image formation model will only be a function of the geometry of the object. This simplifies the shape estimation problem but is generally not an accurate description of real-world objects. We develop a new shape estimation framework that deals with interreflections and can be used on objects with more general reflectance function.

3.2 Problem statement

Given a scene, we first discretize the scene into N points with 3D locations denoted as $\{\mathbf{v}_1, \dots, \mathbf{v}_N\}$, in a right-handed camera-centric coordinate system. The principal point of the camera is at the origin and its optical axis is aligned to the positive z-axis. The 3D location of each scene point can, hence, be expressed as $\mathbf{v}_k = d_k \mathbf{i}_k$, where $d_k \geq 0$ is the depth of the scene point and \mathbf{i}_k is the unit-norm vector that provides the orientation of the scene point from the viewpoint of the camera.

We can measure the path lengths associated with two-bounce light paths. This information can be captured with a collocated projector-camera system with the transient response observed at a pixel p when a different pixel k is illuminated (with a temporal Dirac), the first peak of the received time profile corresponds to a two-bounce path provided that the two scene points are visible to each other. This is because the two-bounce path is the shortest path linking the patch illuminated by pixel

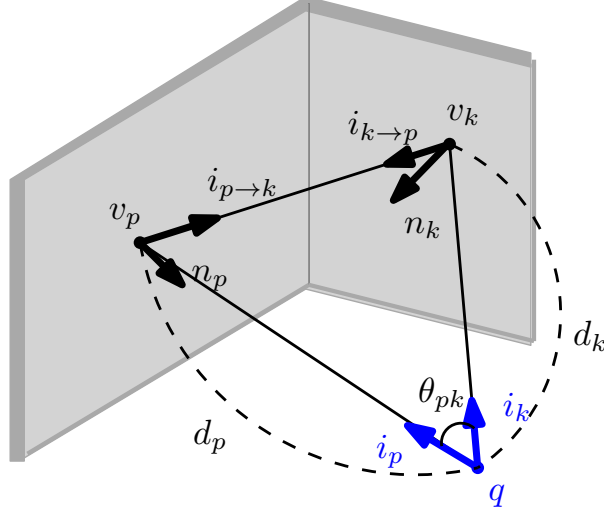


Figure 3.1: **Shape estimation using two-bounce time-of-flight information.** A light path originates from a point \mathbf{q} first bounces off a scene point \mathbf{v}_p then traverses to another scene point \mathbf{v}_k and enters the sensor at \mathbf{q} . The variables labeled in blue are known and form the inputs to the shape estimation problem.

p , and the patch observed by pixel k . Specifically, we measure the time $\tau_{pk}, p, k \in \{1, \dots, N\}$ taken to traverse the two-bounce light path from the origin to scene point p to scene point k and back to the origin (see Figure 3.1). Note that, by Helmholtz’s reciprocity, $\tau_{pk} = \tau_{kp}$. Further, the length associated with these paths is simply given as $c\tau_{pk}$, where c is the speed of light.

The same approach works for non-collocated systems as well. However, the first peak could be from either a single or two-bounce. Single-bounce paths can be identified using: (i) projector pixel p and camera pixel k have to lie on epipolar lines, (ii) the distance measured by triangulation should be consistent with the time-of-flight for the first peak. If these two are simultaneously satisfied, then that corresponds to a single-bounce path between p and k ; otherwise, it is most likely to be a two-bounce light path.

Given the path lengths associated with a collection of two-bounce light paths, our goal is to estimate the scene depths $\{d_1, \dots, d_N\}$. The main advantage of using the path length information is that we only need to consider geometric constraints. The shape estimation problem can be free from estimating the complicated material reflectance function.

3.3 Unique shape recovery analysis with two-bounce light paths

To better describe the depth estimation problem, we characterize all the information available using a graph, called the two-bounce light-path graph. For simplicity, all light paths discussed in this chapter are two bounces. We use the light-path graph as shorthand for the two-bounce light-path graph.

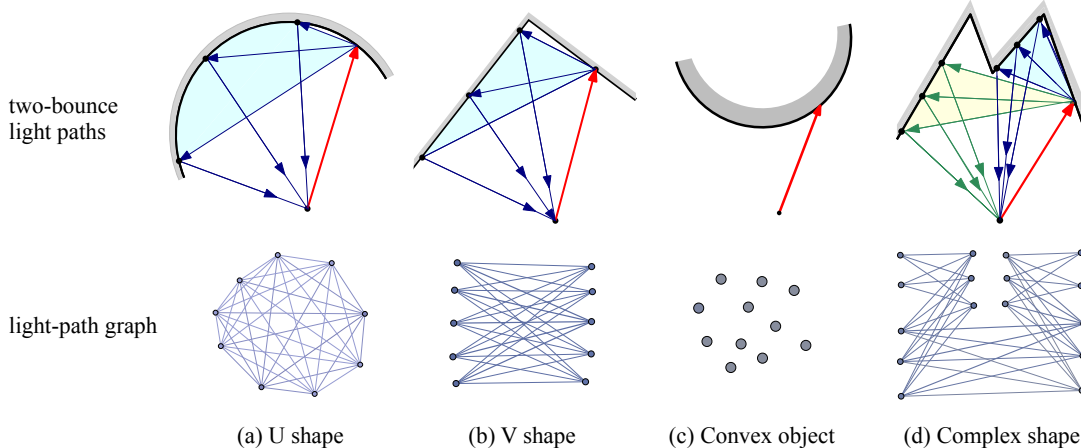


Figure 3.2: **Sample light-path graph for different geometry.** Top row: observable two-bounce light paths of different objects. Bottom row: observable light paths can be represented by a light-path graph.

Light-path graph. We construct a graph $G = (V, E)$, where the vertex set $V = \{1, \dots, N\}$ corresponds to the scene points. An edge (p, k) suggests that the path length associated with the two-bounce light path involving p and k can be observed. Note that, by Helmholtz reciprocity, the graph is non-directional, i.e., the observations associated with the edge (a, b) are identical to that of (b, a) . We refer to the graph G as the light-path graph. It is illustrative to consider the topology of light-path graphs associated with a few classes of scenes (see Figure 3.2).

- *Convex objects.* Convex objects have no two-bounce light paths. Hence, the graph G is simply a collection of vertices with no edges (see Figure 3.2 (c)).
- *Concave objects.* For a generic concave shape — for example, a cup — a two-bounce light path can exist between any two scene points. Hence, in the general setting, the graph is fully-connected or a clique (see Figure 3.2 (a)). However, in practice, in addition to the shape, the reflectance of the scene also plays an important role in determining the topology. For example, for a concave shape with mirror reflectance, the topology of G reduces to isolated edges, i.e., if an edge (p, k) exists, then there are no other edges involving the vertices p and k .
- *V-grooves.* For a generic V-groove, two-bounce light paths can only exist between scene points that belong to opposite sides of the groove. Hence, we can divide the vertex set V into two non-intersecting sub-sets V_1 and V_2 such that edges can only occur between a vertex in V_1 and a vertex in V_2 . Thus V-grooves result in bipartite light-path graphs (see Figure 3.2 (b)).

The light-path graph provides a succinct characterization of all available inputs and its topology is central to the uniqueness of shape recovery.

Uniqueness of shape recovery. Given a light-path graph (or equivalently, a collection of two-bounce light paths and their lengths), can there be multiple shapes that satisfy the path length constraints? The following theorem provides a sufficient condition for uniqueness of shape given a light-path graph.

Theorem 3. *If each connected component¹ of the light-path graph G contains either a cycle with an odd number of vertices or two cycles with an even number of vertices, then the depth of all vertices can be determined uniquely from the available ToF of two-bounce light paths.*

The rest of this section is devoted to the proof of Theorem 3. Note that this result is for the noiseless case where the two-bounce light path lengths are known exactly.

Encoding of depths in two-bounce path length. We now derive the relationships between the depths of scene points and the length of a two-bounce light path. Consider the two-bounce light path associated with two scene points \mathbf{v}_p and \mathbf{v}_k with unknown depths d_p and d_k , and known directions \mathbf{i}_p and \mathbf{i}_k , respectively (see Figure 3.1). From the law of cosines, the total light-path length is given as

$$c\tau_{pk} = d_p + d_k + \sqrt{d_p^2 + d_k^2 - 2d_p d_k \cos \theta_{pk}}. \quad (3.1)$$

Squaring and rearranging (3.1), we can find the following constraint on d_p and d_k :

$$2d_p d_k (-\cos \theta_{pk} - 1) + 2c\tau_{pk}(d_p + d_k) = c^2\tau_{pk}^2. \quad (3.2)$$

Therefore, d_k can be expressed as a function of d_p as

$$d_k = \frac{c^2\tau_{pk}^2 - 2c\tau_{pk}d_p}{2d_p(-\cos \theta_{pk} - 1) + 2c\tau_{pk}} = f_{pk}(d_p), \quad (3.3)$$

$$d_k = \frac{c\tau_{pk}(c\tau_{pk} - 2d_p)}{2(c\tau_{pk} - d_p(1 + \cos \theta_{pk}))}.$$

Since the light path of interest undergoes two bounces, the total path length $c\tau_{pk}$ will be larger than $2d_p$ or $2d_k$. That is, we can bound d_p using

$$d_p < \frac{c \min_{k \neq p}(\tau_{pk})}{2}. \quad (3.4)$$

From (3.4), both the numerator and the denominator of the right-hand side of the expression in (3.3) is positive, so d_k is guaranteed to be positive. Also, the first derivative of $f_{pk}(d_p)$ defined in (3.3) is:

$$f'_{pk}(d_p) = \frac{2c^2\tau_{pk}^2(\cos \theta_{pk} - 1)}{[2c\tau_{pk} - 2d_p(1 + \cos \theta_{pk})]^2}.$$

Since we only consider two-bounce light path, where $p \neq k$, and $\cos \theta_{pk} \neq 1$, we can conclude that

$$f'_{pk}(d_p) < 0. \quad (3.5)$$

Thus $f_{pk}(\cdot)$ is a strictly decreasing function for all feasible d_p .

There are three important consequences to the derivation above. First, given d_p , the value of d_k is uniquely known. Second, f_{pk} is a fractional linear transformation (FLT). Note that the

¹A connected component of a graph is defined as a maximally connected sub-graph.

composition of multiple FLT's remains fractional linear. Third, there are infinite pairs of (d_k, d_p) that satisfy the constraint in (3.3) that relates the two-bounce path length of two individual points. Hence, we cannot solve the depths of two scene points by considering a single two-bounce light path connecting them. Further, it can be shown that the ambiguity inherent in (3.3) cannot be resolved if we consider a light-path graph whose topology is a tree since the relationship between depths is an FLT. This motivates us to explicitly consider light-path graph topologies that include cycles, with the hope that cycles in the graph will help resolve these shape ambiguities. The following Lemma provides a concrete result along this direction.

Lemma 1. *There are at most two solutions for a connected light-path graph whose topology is a single cycle.*

Proof. Since the light-path graph is connected and has a single cycle, all vertices must be part of the cycle. We assume that the cycle in consideration is given by the following edges: (q_1, q_2) , $(q_2, q_3), \dots, (q_{N-1}, q_N)$, and (q_N, q_1) . By applying (3.3) to the edges in succession, we can obtain the following relationship:

$$d_{q_1} = f_{q_N q_1}(\dots f_{q_2, q_3}(f_{q_1 q_2}(d_{q_1}))) = T(d_{q_1}). \quad (3.6)$$

Given that each $f_{q_i q_j}$ is an FLT, the RHS of (3.6) is also an FLT of the form

$$d_{q_1} = \frac{C_1 + C_2 d_{q_1}}{C_3 + C_4 d_{q_1}}, \quad (3.7)$$

where C_1, C_2, C_3 , and C_4 are dependent on the individual two-bounce path lengths. We can rearrange (3.7) to obtain a second-order polynomial equation in d_{q_1} , which has two roots and hence, two potential solutions for d_{q_1} . For each solution, we can estimate the depths of all other vertices uniquely via (3.3). \square

The implications of Lemma 1 are promising since it restricts the solution space associated with a cycle. Further, while there are two potential solutions, it is entirely possible that one of them is infeasible or both solutions are exactly the same. We show that this is indeed the case for cycles with an odd number of vertices.

Proposition 1. *There is exactly one solution for a light-path graph whose topology is a single cycle with odd number of vertices.*

Proof. We prove the uniqueness of the solution by contradiction. Suppose there are two feasible solutions for d_{q_1} from (3.7), $d^{(1)}$ and $d^{(2)}$ with $d^{(1)} < d^{(2)}$. That is,

$$T(d^{(1)}) = d^{(1)}, \quad T(d^{(2)}) = d^{(2)}. \quad (3.8)$$

By chain rule, the first derivative of $T(\cdot)$ is

$$T'(d) = f'_{q_N q_1}(\dots f_{q_2, q_3}(f_{q_1 q_2}(d))) \dots f'_{q_1 q_2}(d).$$

Since the cycle contains an odd number of edges and $f'(d) < 0$ from (3.5), $T'(d) < 0$. Hence, $T(d_p)$ is a strictly decreasing function. Recall that we assumed $d^{(1)} < d^{(2)}$, thus,

$$T(d^{(2)}) < T(d^{(1)}) = d_p^{(1)} < d_p^{(2)}. \quad (3.9)$$

We see a contradiction in (3.8) and (3.9). Therefore, $d^{(1)} = d^{(2)}$ and hence, there is only one feasible solution. \square

Proposition 2. *For a connected light-path graph consisting of two intersecting cycles, each with an even number of vertices, then the depth of all vertices can be uniquely recovered.*

Proof. From Lemma 1, we can find the two potential depths for each of the vertices on the first cycle. Similarly, we can obtain two potential depths for each of the vertices for the second cycle. Since the cycles are intersecting, we can intersect the solutions at a vertex that belongs to both cycles and resolve the ambiguity. \square

As an example, consider the graph in Figure 3.3, the depth of all scene patches can be uniquely recovered. We can first find the two solutions of v_3 , $\{a_1^3, b_1^3\}$ by considering cycle $\{v_1, v_2, v_3, v_4\}$. Similarly, from cycle $\{v_3, v_5, v_6, v_7\}$, we find $\{a_2^3, b_2^3\}$. One element from each set will be consistent, which is the depth estimation for v_3 . After finding d_3 , we can propagate the depth estimation to all vertices in the graph.

Proof of Theorem 3. We now have all the components for the proof of Theorem 3. From Lemma 1, the presence of a cycle in the light-path graph limits the solution space to two. We further prove in Proposition 1 that for a cycle with an odd number of vertices, only one of the solution is feasible. In Proposition 2, when considering multiple cycles, even though each cycle gives two solutions, we can find the unique solution by intersecting the solution across cycles. Therefore, Theorem 3 provides sufficient conditions for shape recovery from light-path graph topology.

Implications of Theorem 3. The sufficient conditions in Theorem 3 only depend on the topology of the light-path graph but not the actual light-path lengths. The topology of the light-path graph depends only on the existence of two-bounce light paths between surface patches, which is completely determined by the geometry of the scene. Hence, Theorem 3 infers uniqueness of shape estimates for a broad class of scenes.

We introduce an algorithm to recover shape from two-bounce light paths. Our goal is in leveraging information over the entire light-path graph to robustly estimate the depth of the scene in the presence of noise corrupting the measured path lengths.

Algorithm 2 (Multi-cycle clustering algorithm.).

One of the main results is that there are at most two depth estimates associated with a cycle in the light-path graph. With the presence of noise, the depth estimation from different sub-graphs will be different. Therefore, we proposed a clustering method to find a solution that is most consistent across different cycles.

We randomly select Q cycles from the light-path graph. By Lemma 1, each cycle provides us with two candidate depths for each of its vertices. We denote the candidate depths for vertex k from cycle q as a_q^k and b_q^k . Repeating this process for each of the selected cycles, we have $2Q$ estimates

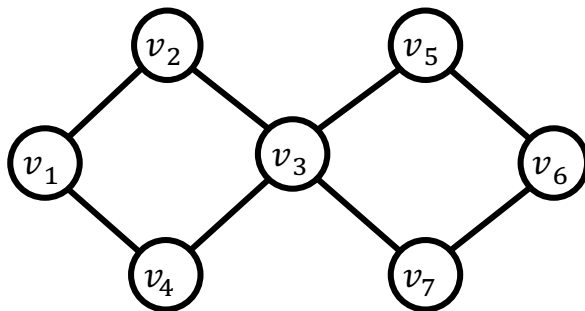


Figure 3.3: **Sample light-path graph to explain Proposition 2.** We form one set of estimation from cycle $\{v_1, v_2, v_3, v_4\}$ and the other set from cycle $\{v_3, v_5, v_6, v_7\}$. There will only be one common estimate for d_3 , which is the true depth of vertex v_3 .

for each vertex. In the absence of noise, Q out of these $2Q$ solutions (one out of 2 solutions from each cycle containing the vertex) would be identical and that would indicate the correct solution for the depth of that vertex. In the presence of noise, the $2Q$ solutions are all potentially different and hence, we need an alternative technique to identify the correct solution.

Let $\mathbf{x} = [x_1, \dots, x_Q]$ be a Q -dimensional indicator vector, where $x_q = 1$ implies that, in cycle q , the candidates in $\{a_q^k\}$ are selected as the estimate and $x_q = -1$ implies that the candidates $\{b_q^k\}$ are selected. Therefore, the depth estimate for vertex k provided by cycle q is

$$d_q^k = \frac{1 + x_q}{2} a_q^k + \frac{1 - x_q}{2} b_q^k.$$

We now minimize the variance of the depth estimates:

$$\min_{\mathbf{x} \in \{-1, 1\}^Q} \sum_{k=1}^N \left[\frac{1}{Q} \sum_{q=1}^Q (d_q^k)^2 - \left(\frac{1}{Q} \sum_{q=1}^Q d_q^k \right)^2 \right]$$

This is a combinatorial problem due to the feasible set being $\mathbf{x} \in \{-1, 1\}^Q$. If we relax this constraint to let \mathbf{x} take real values, and observing that the objective is quadratic in \mathbf{x} , we can obtain a closed form solution for \mathbf{x} . By thresholding this solution at zero, we get our estimator of the indicator vector $\hat{\mathbf{x}}$. We use the mean of candidate depths at each vertex selected by $\hat{\mathbf{x}}$ as our estimate.

3.4 Experiments

Synthetic experiment – Gaussian noise. In Figure 3.4, we show the recovery results on a 1D slice of an object for varying amount of additive Gaussian noise on the measured path lengths. We compare single-bounce estimates and the multi-cycle clustering. We use $Q = 100$ cycles for the multi-cycle method. We can see that multi-cycle clustering outperforms single-bounce technique by a large margin, in part, due to the availability of a larger number of light paths.

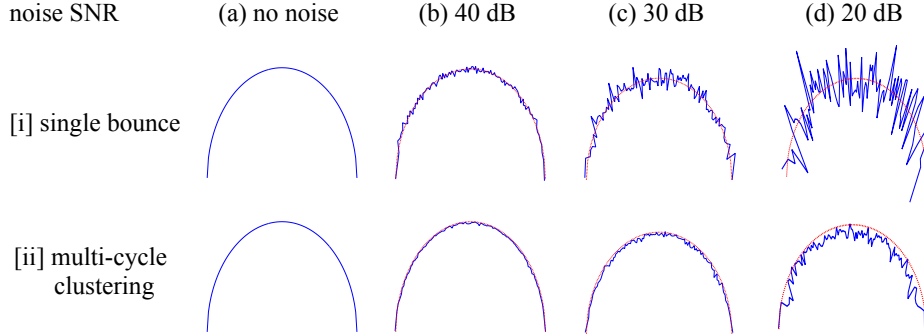


Figure 3.4: **2D shape recovery results.** We perturbed the input using different levels of noise. The blue line shows our reconstructed surface and the red dotted line is the ground truth.

Synthetic experiment – Physically accurate rendering. While we explain how a light-path graph is constructed, we mention that the reflectance of the object will impact where a two-bounce light path is observable or not. We use a time-of-flight extension [103] of physically accurate rendering software, Mitsuba [56], to render the light transients. We illuminate the scene with collimated beam emitter at 144 points on a regular grid. We set the camera to have a spatial resolution of 128×128 . We use the first peak above a set threshold in the light transient in each pixel as the two-bounce light path. We use 25 cycles for the multi-cycle clustering algorithm.

As shown in Figure 3.5, we test the proposed algorithm using 5 concave objects with 5 reflectance. Four of the reflectance are from the MERL BRDF database [86] and rendered using the Mitsuba plugin developed by Dupuy [26]. An example of the rendered transient is shown in Figure 3.6. Notice that the rendered transients consist of one-bounce, two-bounce, and also higher-order bounce light paths. During the reconstruction, we do not retain the bounce information. We block the one-bounce light by placing a mask on the neighborhood pixels of the illuminated direction. For the light path that is higher-order bounce and its intensity being above a threshold, we assume it to be measurable and they are a source of noise in the light-path graph.

We visualize the recovered results of our proposed multi-cycle clustering algorithm in Figure 3.7. We can see that the algorithm works well with different shapes and reflectances.

We test the algorithm by altering the threshold to determine the shortest light path. Exp 1 - 4 corresponds to a threshold of 10^{-5} , 2×10^{-5} , 5×10^{-5} , and 10×10^{-5} , respectively. We show two metrics to determine the quality of the reconstruction. First, the root mean squared error (rmse), $\sqrt{\frac{1}{N} \sum (d_i - d_i^{\text{gt}})^2}$. Another metric is coverage. Since we only illuminate on sparse points sequentially, there are some parts of the scene that never receive interreflected light. We cannot possibly recover the depth of these parts from two-bounce light paths. We show the ratio of the number of recovered points and the number of ground truth points.

As expected, when increasing the threshold, the number of observable light path decreases (see Figure 3.8). The spatial extent of interreflection decreases. Therefore, the coverage of recovered points decreases. We visualize the light-path graph of the corner scene in Figure 3.9. We use the adjacency matrix to visualize the light-path graph. When the threshold is large, fewer light paths are observable. We can see that less node in the light-path graph is connected. When the threshold

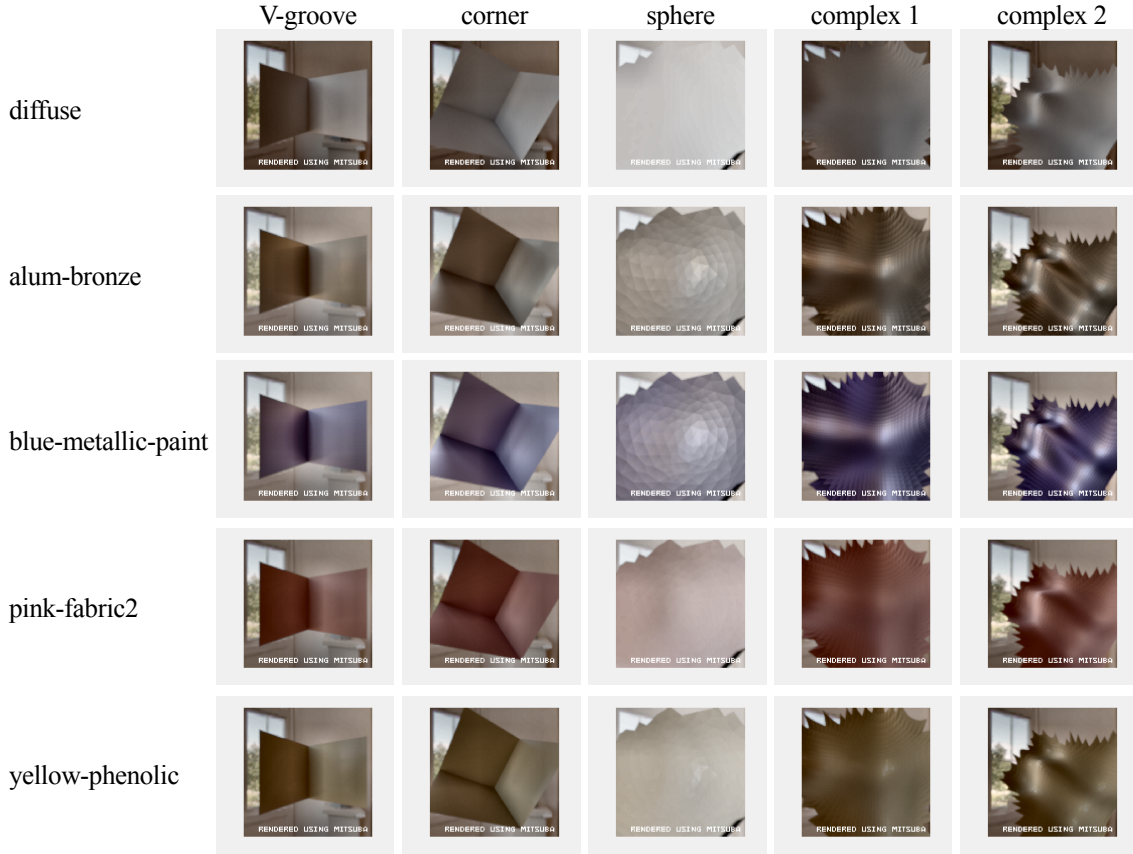


Figure 3.5: **Visualization of scene shapes and materials.** We test 5 object shapes and 5 different materials. We illuminate the scene with an environment map to help visualization.

is large, the sufficient condition may be broken, thus no shape can be recovered.

3.5 Conclusion

While there is a rich body of work associated with all aspects of shape and reflectance estimation from single-bounce light paths, little is known in terms of the capabilities and limitations of two-bounce light paths. Our specific contributions are as follows.

- *Two-bounce shape estimation.* We propose a formulation for the systematic study of shape estimation from two-bounce light paths and develop a graph-based framework to represent the information in such light paths.
- *The uniqueness of shape estimation.* We establish sufficient conditions for uniqueness of shape estimation given only second-bounce light paths. Our sufficient conditions are intricately tied to the topology of the graph that characterizes the available second-bounce light paths.

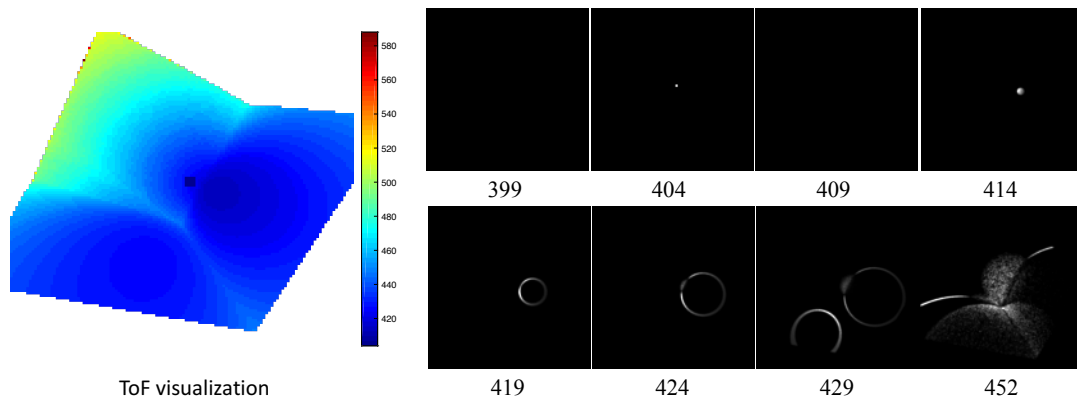


Figure 3.6: **Example of rendered transients for light-path graph determination.** We show several time instants of the rendered light transient. The scene is a corner with diffuse reflectance. Each image is normalized to have a maximum intensity equal to one for clearer visualization.

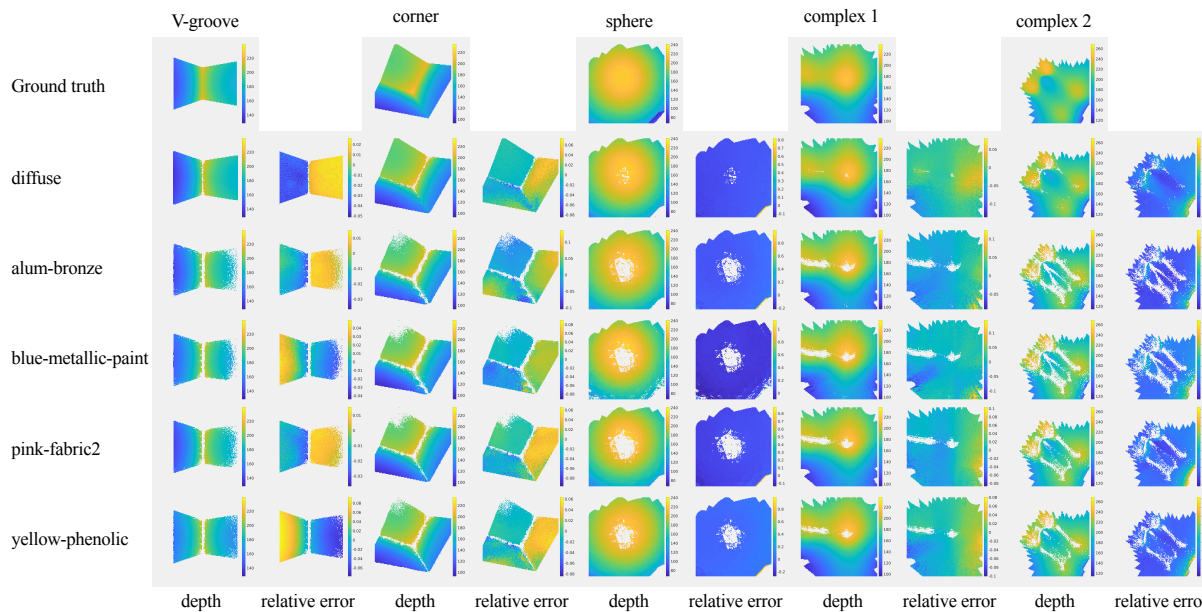


Figure 3.7: **Experiment results for shape from two-bounce light paths.** We show the recovered depth map and the relative error. The threshold is 10^{-5} , corresponding to exp1 in Table 3.1.

- *Algorithms for shape estimation.* We provide novel algorithms that estimate shape given lengths of two-bounce light paths.

By understanding the capacity of two-bounce light paths, we provide a roadmap for holistic methods that exploit the single- as well as multi-bounce light paths.

material	shape	exp1		exp2		exp3		exp4	
		rmse	coverage	rmse	coverage	rmse	coverage	rmse	coverage
diffuse	V-groove	5.1240	0.9837	3.2300	0.9734	3.4546	0.9342	4.3712	0.8501
	corner	3.6318	0.9792	3.7998	0.9632	4.1487	0.9023	4.3174	0.8533
	sphere	5.1934	0.9840	6.3812	0.9631	5.3332	0.9303	5.1421	0.8978
	complex 1	3.7053	0.9977	3.6787	0.9860	3.2650	0.9648	3.6822	0.9464
	complex 2	6.0555	0.9586	6.4231	0.9215	7.0077	0.8452	6.9327	0.7834
alum-bronze	V-groove	2.3667	0.8524	11.3908	0.7257	1.8475	0.5349	10.6206	0.4092
	corner	6.6491	0.8792	5.2783	0.8272	4.6143	0.7299	4.7870	0.6247
	sphere	6.6868	0.9314	6.1487	0.9001	5.9385	0.8404	5.7151	0.7294
	complex 1	3.3316	0.9560	3.4467	0.9292	3.1290	0.8525	3.9681	0.7181
	complex 2	6.5582	0.7756	8.4432	0.6926	8.5096	0.5554	N/A	N/A
blue-metallic-paint	V-groove	3.5384	0.8475	3.4125	0.7523	3.6335	0.6012	13.8015	0.4556
	corner	4.3212	0.8610	4.3815	0.8039	7.4340	0.6941	5.4802	0.5719
	sphere	4.7653	0.9096	3.9462	0.8190	2.0299	0.6827	1.5973	0.5806
	complex 1	3.4295	0.9554	3.2168	0.9286	2.9964	0.8250	3.3878	0.6566
	complex 2	6.1369	0.7726	6.2817	0.6816	6.0146	0.5249	N/A	N/A
pink-fabric2	V-groove	2.5101	0.8786	1.5797	0.7559	9.1745	0.4929	5.0909	0.2763
	corner	4.2419	0.8799	4.7506	0.8252	5.0603	0.7357	8.7573	0.4678
	sphere	5.2816	0.9110	5.7094	0.8742	9.5981	0.3904	16.4886	0.0191
	complex 1	3.6694	0.9580	4.2341	0.9303	4.4055	0.7749	N/A	N/A
	complex 2	8.9125	0.8299	7.3735	0.7474	8.1305	0.5800	N/A	N/A
yellow-phenolic	V-groove	8.9353	0.9346	6.3941	0.8228	2.8697	0.6481	4.3922	0.4282
	corner	3.9170	0.9096	3.9803	0.8554	4.5516	0.7704	6.8140	0.6759
	sphere	6.3892	0.9381	5.6554	0.9061	6.7991	0.8554	6.1285	0.7709
	complex 1	3.6254	0.9690	4.6277	0.9488	3.6384	0.9018	3.4216	0.8212
	complex 2	7.4725	0.8544	6.5767	0.7762	10.0326	0.6609	8.1453	0.5233

Table 3.1: **Alternating image intensity threshold used for measuring two-bounce light paths.** We alter the threshold used to determine whether a two-bounce light path is observable. Exp 1 corresponds to the smallest threshold and Exp 4 corresponds to the largest threshold.

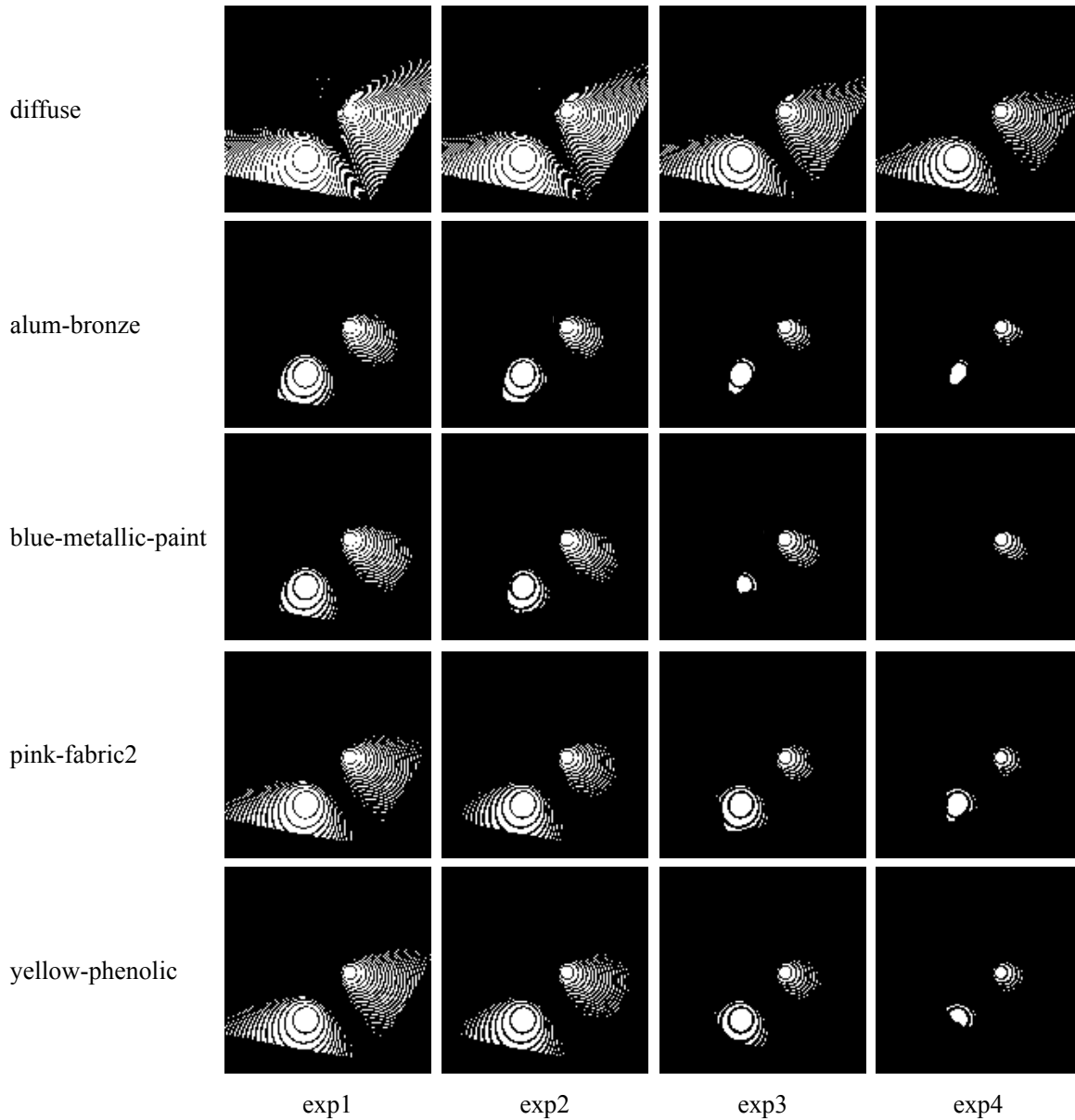


Figure 3.8: **Observable light paths.** We show examples of the observable light paths. The scene is the corner and we illuminate a spot on the scene. The interreflections on the two other walls will be used for shape estimation. Since the threshold used in exp4 is largest, the number of observable two-bounce light paths is smaller.

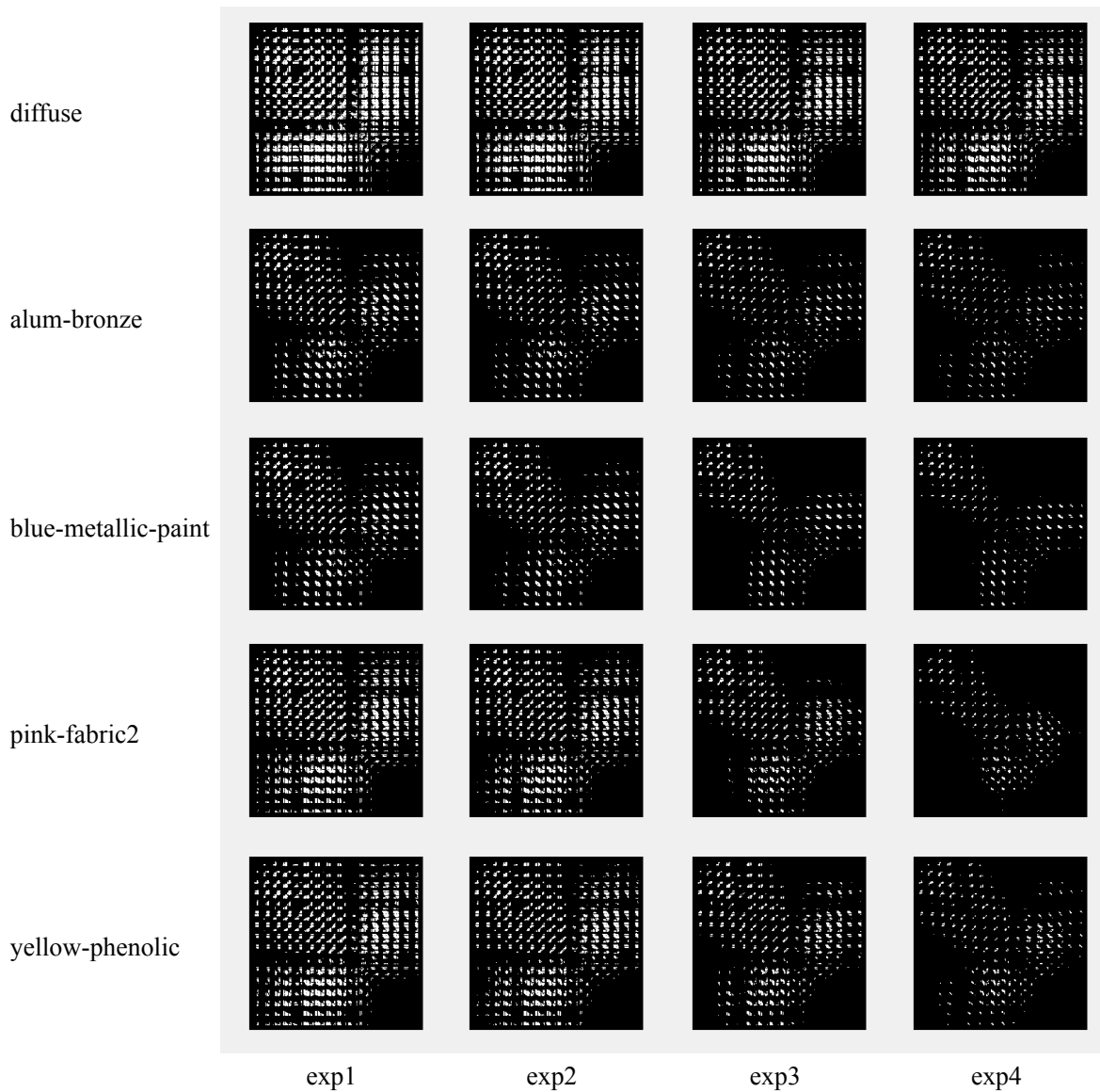


Figure 3.9: **Light path graph visualization.** We show the light-path graph used in the experiments. The scene is the corner. We visualize the light-path graph using adjacency matrices where the $(i-j)$ entry represents the connectivity of vertex i to vertex j .

Chapter 4

Non-line-of-sight imaging — Shape from shortest three-bounce light paths

In Chapter 3, we show how two-bounce light paths are valuable for shape estimation. We can directly utilize the embedded information in the light paths and not just treat multi-bounce light paths as undesirable noises. However, we argue that multi-bounce light paths are not just something that we need to design special algorithms to work around, they actually provide unprecedented potentials for imaging. Non-line-of-sight (NLOS) imaging [69, 135] (see Figure 4.1) is one of the best examples for the use of multi-bounce light paths. NLOS imaging refers to estimation of the shape, texture, and reflectance of scene points that lie *beyond* the field of view of an imaging system. The object of interest is never observed with direct light paths since it is not in the field of view from the observer. The information of the object is only contained in multi-bounce light paths.

There are numerous approaches for estimating the shape of the scene within the field of view of the imaging system. In contrast, NLOS shape estimation is a challenging task requiring capture and analysis of photons that have traveled beyond the line of sight (LOS). This is typically achieved by measuring the so-called 5D light transient transport tensor [95, 110] which captures light propagation — from the LOS scene onto the NLOS scene and back — at ultra-high temporal resolutions (typically, in picoseconds).

We use a pulsed source I_0 and a transient detector s_0 to image a scene that consists of two distinct sets of surfaces: surfaces \mathcal{S}_{LOS} that are visible to both the source and detector, and surfaces $\mathcal{S}_{\text{NLOS}}$ that are occluded from both of them. We assume that there are no surfaces that are neither in \mathcal{S}_{LOS} nor in $\mathcal{S}_{\text{NLOS}}$. We additionally assume that the visible surface \mathcal{S}_{LOS} has Lambertian reflectance.

We use the source to illuminate a point l on the visible surface \mathcal{S}_{LOS} . Likewise, we use the detector to image a point s on \mathcal{S}_{LOS} . We call the points l and s the virtual source and virtual detector, respectively. This terminology stems from the fact that these points effectively act as an isotropic source and detector directly attached to \mathcal{S}_{LOS} , as they redirect light, through a diffuse reflection, from the source to the NLOS scene, and from the NLOS scene to the detector.

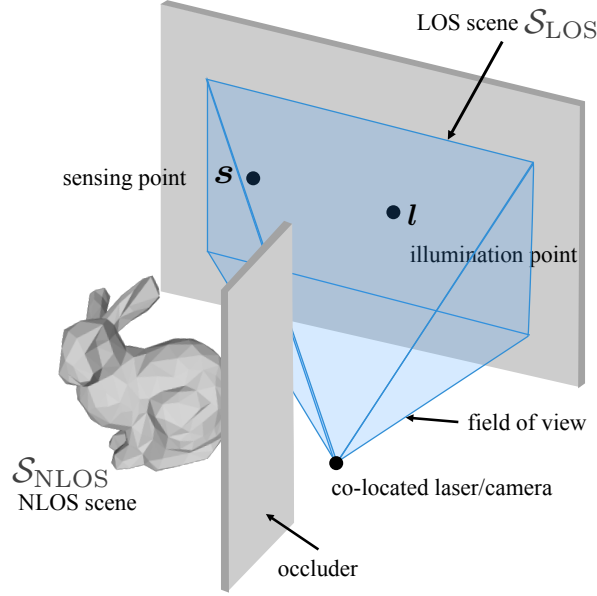


Figure 4.1: **Non-line-of-sight imaging setup.** NLOS imaging uses properties measured in LOS \mathcal{S}_{LOS} to infer the properties of the NLOS scene $\mathcal{S}_{\text{NLOS}}$. We use the first-returning photon associated with illuminating point l and sensing point s to infer the shape of the NLOS scene.

We define $I(t; l, s)$ as the light transport at time t when observing at s and illuminating l with an impulse at $t = 0$. We refer to

$$\{I(t; l, s) \mid \forall l, s \in \mathcal{S}_{\text{LOS}}\}$$

as the 5D light transient tensor since it encodes one degree of freedom along time and two spatial degrees, each, for both illumination and sensing points.

For simplicity of explanation, let's assume that the LOS scene \mathcal{S}_{LOS} is convex; hence no interreflections of photons within LOS. When $l \neq s$, the convexity of \mathcal{S}_{LOS} implies that there are no single- or two-bounce light paths from the laser to the camera; in fact, there are no light paths from the laser to the camera that involve *only* LOS scene points. Hence, a non-zero intensity in the light transient $L(t; l, s)$ encodes properties associated with three- and higher-bounce light paths that include LOS as well as NLOS scene points. This forms the basis of NLOS imaging.

We focus solely on three-bounce light paths from the laser to the camera. We assume that the locations of all LOS scene points are known and, for simplicity, that the laser and camera are co-located at the origin. Given this, a non-zero intensity in $L(t; l, s)$ at $t = t_0$, indicates the presence of three-bounce light path(s) whose length is t_0 ¹ and implies the presence of NLOS scene point(s) x that satisfy

$$\tau(x) = \|x - l\| + \|x - s\| = t_0 - \|l\| - \|s\|, \quad (4.1)$$

¹We treat geometric pathlength and time of flight as equivalent, with the understanding that they relate to each other through the speed of light.

where $\mathbf{x} \in \mathbb{R}^3$. This constrains \mathbf{x} to lie on an ellipsoid whose foci are at \mathbf{l} and \mathbf{s} [15, 135]. Moreover, the intensity $L(t_0; \mathbf{l}, \mathbf{s})$ is an integral that encodes both the shape and the reflectance of all NLOS scene points on the ellipsoid defined in (4.1). To further simplify, it is common to assume that the NLOS scene is Lambertian so that the BRDF is represented by a spatially-varying *scalar* albedo pattern. By parsing through all entries in the entire 5D transient, namely different illumination and sensing points as well as time instants, we can produce a large number of ellipsoidal integral constraints on the NLOS albedo. The albedo is subsequently recovered by solving a complex inverse problem [135].

First-returning photons. By considering how photons traverse the scene, we find out some photons carry more information than others. They are the first-returning photons, which correspond to the shortest light paths. We provide a new formulation for NLOS shape recovery that avoids solving an inverse problem altogether. Specifically, instead of considering all NLOS light paths, we only focus on the path length associated with the first-returning photon, defined as follows:

Definition. The *first-returning photon* at an LOS point \mathbf{s} , when we illuminate \mathbf{l} , is the photon that traverses the shortest three-bounce light path through \mathbf{l} and \mathbf{s} , i.e., it is the first photon that we observe at \mathbf{s} while illuminating \mathbf{l} .

The time of flight (ToF) of the first-returning photon is the smallest time instant t_0 such that $L(t_0; \mathbf{l}, \mathbf{s})$ is non-zero. Its path length, $\delta(\mathbf{l}, \mathbf{s}) = t_0$, is given as

$$\delta(\mathbf{l}, \mathbf{s}) = \|\mathbf{l}\| + \|\mathbf{s}\| + \min_{\mathbf{x} \in \mathcal{S}_{\text{NLOS}}} \|\mathbf{x} - \mathbf{l}\| + \|\mathbf{x} - \mathbf{s}\|. \quad (4.2)$$

A key observation is that the shortest path is often unique and determined by the position of a *single* NLOS scene point. This greatly simplifies the shape estimation problem, since we do not need to unmix the contributions of many NLOS scene points. Also, we explore the geometric constraints of first-returning photons. By using only a fraction of the information from the 5D transient, i.e., the first-returning photons, we developed simple algorithms to locate hidden points and their associated normals. Instead of solving an ellipsoidal tomography problem to recover the albedo, we use the shortest light-path property of first-returning photons to recover the 3D points on the hidden object.

In this chapter, we show our shape recovery technique based on shortest light paths [126, 127].

4.1 Related work

We discuss some works in non-line-of-sight imaging. NLOS imaging refers to the broad problem of reconstructing properties of scenes that are normally occluded from a sensor. Even though interest in this problem dates back several decades [31], it has recently attracted increased attention within computer vision and graphics, following two seminal papers [69, 135] demonstrating the ability to reconstruct shape in the *looking around the corner* setting (Figure 4.1). Ever since, many efforts have been made in this field. Related literature roughly falls into two categories. First, algorithms that make use of time-resolved measurements to form constraints of the NLOS objects [59]. Second, algorithms that use image or video intensities to reason the NLOS scene. An overview of the literature is in Table 4.1.

Measurement	Methods	Reference
Time resolved	linear systems	Heide <i>et al.</i> [50], Heide <i>et al.</i> [49]
	backprojection	Velten <i>et al.</i> [135], Buttafava <i>et al.</i> [15], Manna <i>et al.</i> [75], Arellano <i>et al.</i> [5]
	surface prior	Kadambi <i>et al.</i> [61], Gariepy <i>et al.</i> [34], Pediredla <i>et al.</i> [102]
	closed-form solution	O’Toole <i>et al.</i> [96]
	first-returning photon	Chapter 4 [126, 127]
	surface optimization	Chapter 5 [128]
Intensities	NLOS object prior	Klein <i>et al.</i> [71], Thrampoulidis <i>et al.</i> [124], Xu <i>et al.</i> [144]
	video sequence	Bouman <i>et al.</i> [12]
	speckle	Katz <i>et al.</i> [66], Katz <i>et al.</i> [65], Smith <i>et al.</i> [118], Bertolotti <i>et al.</i> [8], Boger-Lombard and Katz [9]

Table 4.1: NLOS scene recovery related works.

Time-resolved measurements provide ellipsoidal constraints on the albedo of the NLOS object [15, 135]. The forward model for the 5D light transport tensor is linear with respect to the albedo and geometry of the hidden object. By neglecting the effect of the surface normal and voxelizing the space, the spatial-varying albedo can be recovered. Based on this image formation model, researchers focus on the following four directions: solving a linear inverse problem, speeding up backprojection, introducing NLOS models, and finding a closed-form solution for the linear inverse problem. Heide *et al.* [50] demonstrate NLOS shape reconstruction directly from a PMD’s readout by relating the measurement to linear systems. More recently, Heide *et al.* [49] take into account of visibility issues and improve the formation model. However, depending on the number of voxels and the number of available measurements, the computation time for the inverse problem can be high. One way to avoid directly solving the inverse system is to use backprojection to find voxels that are most related to the measurements. Manna *et al.* [75] propose an iterative solution that further updates the result of backprojection. By rendering the light transient of the estimated hidden surface and compare against the measurement, they use the difference to further refine the estimation. Arellano *et al.* [5] speed up backprojection by only considering voxels where multiple ellipsoids intersect, this decreases the computation needs. However, the resolution of the recovered scene depends on the number of ellipsoidal constraints. Another common approach to decrease the complexity of solving the inverse problem is to introduce surface priors. Kadambi *et al.* [61] use signal sparsity to recover distinct scene points. Gariepy *et al.* [34] approximate the hidden scene with isolated scene points. Each scene point becomes a light source generating a spherical wavefront. The source can be located by intersecting multiple ToF measurements. Pediredla *et al.* [102] estimate planar scenes by comparing SPAD readout with rendered results. More recently, O’Toole *et al.* [96] show the transient measurements can be interpreted as 3D convolution by change of variable in the image formation model and provide a closed-form solution for the albedo of each voxel.

In this chapter, contrary to all methods based on time-resolved measurement, we propose using first-returning photons and derive geometric constraints based on shortest paths. We provide algorithms that only require the ToF of first-returning photons. In the next chapter, we discuss how most methods are making physically inaccurate simplification for the image formation model and show a surface optimization framework using time-resolved measurements to estimate shape or refine the results we gathered from algorithms discussed in this chapter.

NLOS imaging can be achieved without time-resolved measurement. Klein *et al.* [71] perform NLOS pose estimation using just image intensities. In this case, the number of unknowns is smaller and image intensities are sufficient for estimating the hidden object. By adding additional known occluder in the NLOS scene, Thrampoulidis *et al.* [124] and Xu *et al.* [144] show that image formation model is easier to invert since the measurement only contains the contribution from a subset of the unknown scene. Bouman *et al.* [12] use the fact that visible parts around the corner can see different portions of the NLOS objects and recover the motion of hidden objects by taking a series of videos and finding intensity variation across time. Katz *et al.* [65] utilize the fact that the autocorrelation of the image with and without scattering is the same. The hidden object can be recovered with the acquired autocorrelation image. Smith *et al.* [118] use the fact that the motion of hidden objects will reflect on speckle pattern motion to track the object’s movement. By assuming the hidden source is a light source emitting continuous broadband noise, Boger-Lombard and Katz [9] propose a passive optical architecture to localize the hidden source. By using a two-aperture mask, a pixel of the camera can measure the temporal cross-correlation via interferometry of two parts of the LOS wall. Based on the acquired cross-correlation, the hidden source can be subsequently localized.

4.2 Problem statement

We assume an imaging system consisting of a pulsed laser and a ToF camera. We image the NLOS object through a diffuse LOS scene. Given two LOS scene points $\mathbf{l}, \mathbf{s} \in \mathcal{S}_{\text{LOS}}$, we assume that the setup can measure the path length associated with the first-returning photon at \mathbf{s} when we illuminate \mathbf{l} with an impulse. Here we only consider first-returning photons from three-bounce light paths.

Since the first-returning photon traverses the shortest three-bounce light path involving \mathbf{l} and \mathbf{s} , its path length $\delta(\mathbf{l}, \mathbf{s})$ is given in (4.2). Given a known LOS scene, we can deduct from $\delta(\mathbf{l}, \mathbf{s})$ the distance of the laser and camera to illumination point \mathbf{l} and sensing point \mathbf{s} , respectively. This provides us with the length of the shortest path from \mathbf{l} to \mathbf{s} via the NLOS scene,

$$d(\mathbf{l}, \mathbf{s}) = \min_{\mathbf{x} \in \mathcal{S}_{\text{NLOS}}} \|\mathbf{x} - \mathbf{l}\| + \|\mathbf{x} - \mathbf{s}\|. \quad (4.3)$$

Our goal is to identify points belonging to the NLOS scene $\mathcal{S}_{\text{NLOS}}$ given a collection of shortest path lengths:

$$\{d(\mathbf{l}, \mathbf{s}) \mid \mathbf{l}, \mathbf{s} \in \mathcal{S}_{\text{LOS}}\}.$$

4.3 Acquisition of first-returning photons

In this section, we first detail how first-returning photons are acquired with a SPAD system and provide a theoretical analysis of the estimation of the first-returning photon through this process. Finally, we highlight the advantages of a shape recovery framework that uses first-returning photons. We show both theoretical analysis and realistic experimental results to support the benefits of the first-returning photons.

Light transient estimation using a SPAD. Here we outline the process to estimate the light transient using a SPAD. We only highlight SPAD operations that are related to our analysis and defer to [15] for a more comprehensive description of acquiring light transients with a SPAD.

The probability of receiving a photon at a given time instant is determined by the intensity of the light transient at the time instant. For NLOS imaging, the light path undergoes three scattering events and, as a consequence, the probability of receiving a photon is low. When a photon is incident, the SPAD saturates and records the time-stamp of the event. We can repeat this process multiple times by sending an impulse train from the laser. By counting the number of photons arriving in each time bin, we measure a photon distribution that is determined by the light transient. A larger integration time means repeating the photon collection process for a longer time; thus more photons are collected for the estimation of photon distribution.

It is worth noting that the photon distribution is not directly proportional to the light transient, since there are nonlinear processes in the SPAD image formation model that includes time jitter, pile-up, dead time, internal noise, ambient noise, and other operational effects [19, 51]. For simplicity of analysis, we ignore these non-linearities in the SPAD operation and approximate the photon distribution collected by SPAD to be proportional to the light transient. After the light transient is measured, we extract the first-returning photons as the first time instant when the photon count increases drastically.

Acquisition advantages of the first-returning photon. In the following, we provide both theoretical analysis and experimental validation of the advantages of first-returning photons.

First, we consider a simplified SPAD image formation model to analyze the statistical characteristic of the ToF of the first-returning photons. We disregard any noise and use the first non-zero time instant in the light transient as the ToF of the first-returning photon. Let the arrival time of the first N recorded photon, T_1, T_2, \dots, T_N , be independent and identically distributed random variables with probability density function (PDF), $f(t) = g(t)u(t - t_{fp})$. Here, we form $f(t)$ as the product of two functions to highlight that the ToF of the true first-returning photon to be t_{fp} , and that we don't receive photons before t_{fp} .

An estimate of the ToF of first-returning photon is given as

$$\hat{T}_{fp} = \min\{T_1, T_2, \dots, T_N\}.$$

The cumulative distribution function (CDF) of \widehat{T}_{fp} is given as

$$\begin{aligned}
F_{\widehat{T}_{\text{fp}}}(t) &= P(\widehat{T}_{\text{fp}} \leq t) \\
&= P(\min\{T_1, T_2, \dots, T_N\} \leq t) \\
&= 1 - P(T_1 > t)P(T_2 > t) \cdots P(T_N > t) \\
&= 1 - [1 - P(T_1 \leq t)]^N.
\end{aligned}$$

Hence, the PDF of the ToF of the first-returning photon is

$$\begin{aligned}
p_{\widehat{T}_{\text{fp}}}(t) &= \frac{d}{dt} F_{\widehat{T}_{\text{fp}}}(t) \\
&= \frac{d}{dt} \{1 - [1 - P(T_1 \leq t)]^N\} \\
&= N[1 - P(T_1 \leq t)]^{N-1} f(t).
\end{aligned}$$

We can show the number of photons needed to decrease the probability of the first-returning photon being ϵ away from the true first-returning photon:

$$\begin{aligned}
P(\widehat{T}_{\text{fp}} > t_{\text{fp}} + \epsilon) &= 1 - P(\widehat{T}_{\text{fp}} \leq t_{\text{fp}} + \epsilon) \\
&= [1 - P(T_1 \leq t_{\text{fp}} + \epsilon)]^N
\end{aligned} \tag{4.4}$$

Even though the exact probability is dependent on the specific PDF of the arrival time of each photon, an important implication of (4.4) is that the probability exponentially decays with respect to the number of received photon N . To give a rough idea of the number of photons needed, we use a sample scene that consists of a planar patch and render the light transient to show related quantities (see Figure 4.2).

$$P(\widehat{T}_{\text{fp}} > t_{\text{fp}} + 33.3) = 0.05 \geq [1 - P(T_1 \leq t_{\text{fp}} + 33.3)]^N$$

With only $N = 103$ photons, the probability of getting an estimate of first-returning photon being 33.3 picoseconds away from ground truth is less than 0.05.

From several realistic light transients collected by SPAD, we observe that light transient can be approximated with an exponential function. We discuss this special case of PDF being an exponential distribution, $f(t) = ae^{-a(t-t_{\text{fp}})}u(t-t_{\text{fp}})$, where t_{fp} is the true ToF of the first-returning photon. The CDF of first-returning photon estimate becomes:

$$F_{\widehat{T}_{\text{fp}}}(t) = [1 - e^{-aN(t-t_{\text{fp}})}]u(t-t_{\text{fp}})$$

Hence,

$$P(\widehat{T}_{\text{fp}} > t_{\text{fp}} + \epsilon) = e^{-aN\epsilon}.$$

The probability of the estimator being ϵ away from the true value exponentially decreases with N . The probability is highly concentrated at $\epsilon = 0$.

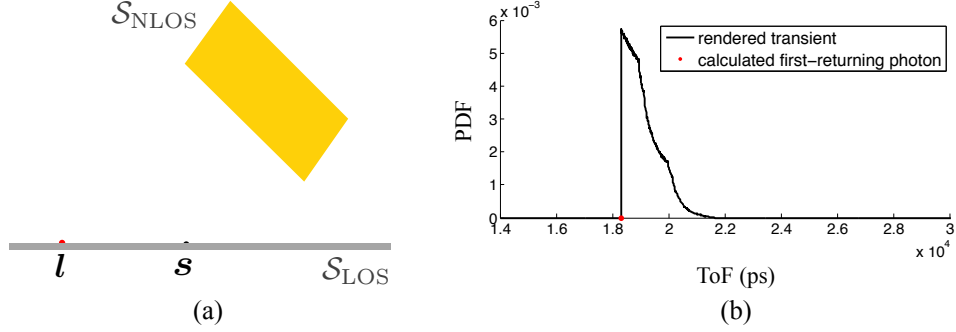


Figure 4.2: **Sample scene and rendered light transient.** To demonstrate the number of photons needed for the estimation of the first-returning photon to be close to the true first-returning photon, we synthesize the light transient of a planar NLOS object by ray tracing 10^8 photons.

We now show several realistic results to demonstrate how first-returning photons are easy to acquire. Most of the previous works in NLOS use the full 5D light transient to solve the inverse ellipsoid tomography problem. This requires a high-quality estimation of the light transient. For SPAD systems, higher SNR can be achieved by a longer integration time (see Figure 4.3). For shorter integration time, the number of photons arriving at each bin varies drastically. However, the ToF of the first-returning photon remains stable even for a short integration time (see Figure 4.4). We further repeat the experiment of estimating the position of the first-returning photon in each peak in the transient multiple times to gather the statistical behavior of the estimation (see Figure 4.5). As shown in Figure 4.3, the SNR of the first-returning photon estimation is significantly better compared to the SNR of the whole transient. That is, the ToF of the first-returning photon can be estimated reliably with short exposure time. This is also consistent with our theoretical derivation in (4.4). Hence, shape reconstruction techniques that rely on first-returning photons will have a significant advantage at data acquisition.

Robustness to material reflectance. In addition to acquisition advantages of the first-returning photon, a shape estimation framework based on first-returning photon is useful to deal with unknown NLOS material reflectance. The ToF of a light path is a function of the geometry of the NLOS object. As shown in Figure 4.6, even though the transient response changes with respect to the reflectance, the ToF of the first-returning photon remains the same.

4.4 Geometry of first-returning photons

In the previous section, we derive the advantages of first-returning photons for NLOS imaging. In the following, we will derive geometric constraints on the NLOS scene given the path length of the first-returning photon when illuminating l and sensing at s . These geometric properties are the keys that enable NLOS shape recovery.

Ellipsoidal constraint. Given $d(l, s)$, as in (4.3), we can deduce that there is at least one NLOS scene point lying on an ellipsoid $E(l, s)$ given as:

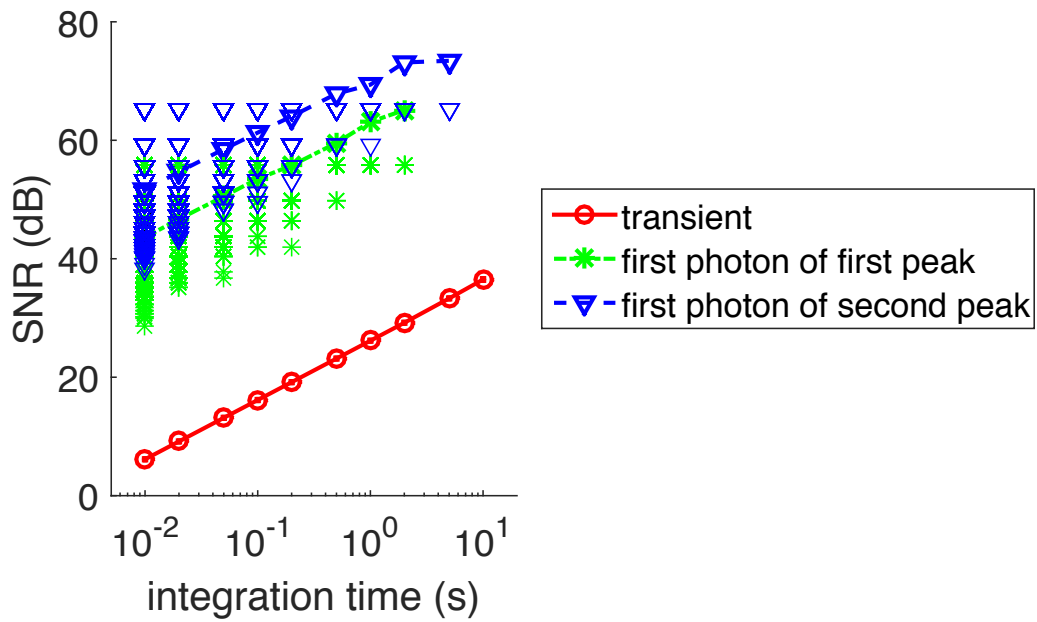


Figure 4.3: **The SNR of acquired transient and first-photon estimation with respect to the integration time.** We use the longest integration time (100 seconds) as the noiseless signal, and compare it with all other shorter integration time signals. We also normalize the height of the transient response so that the total number of laser pulses is the same.

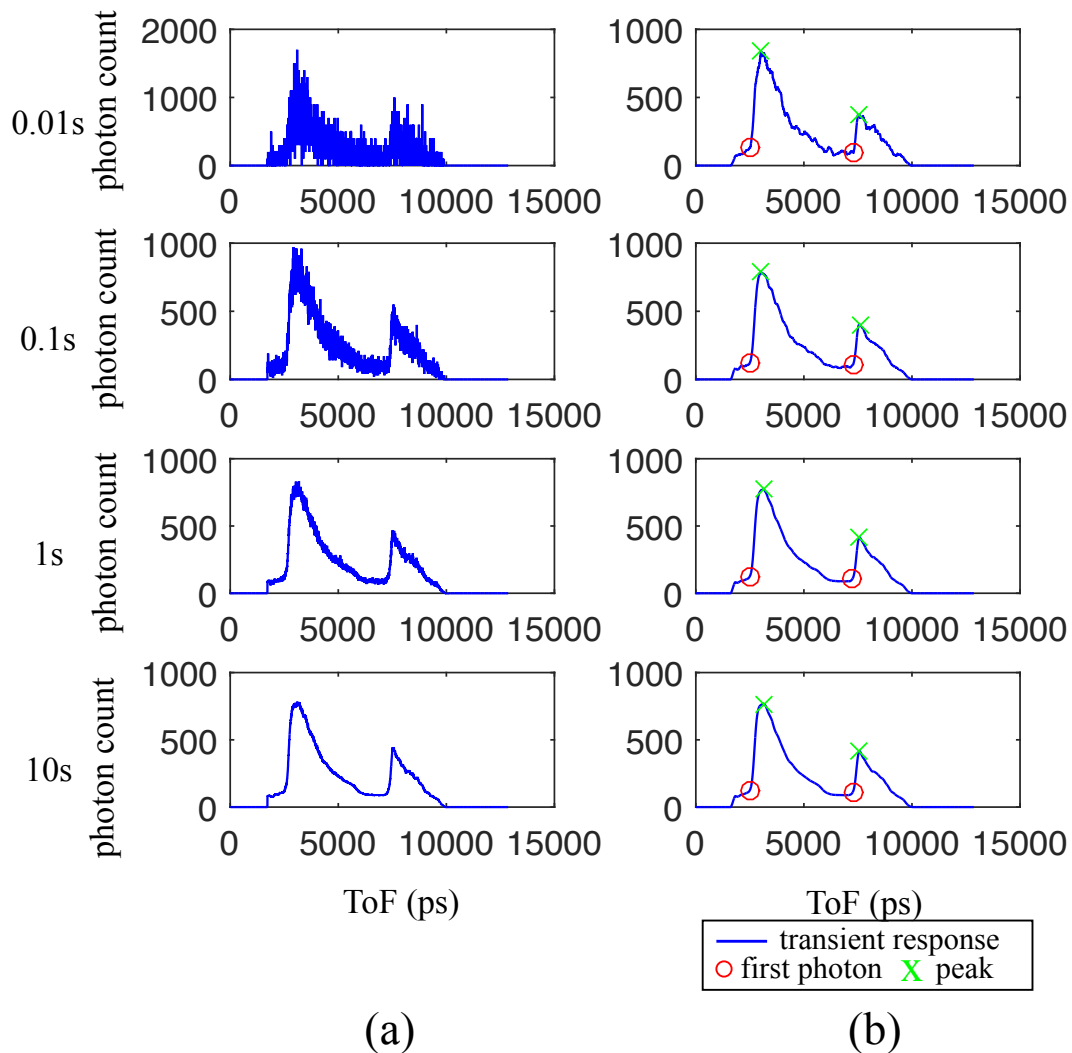
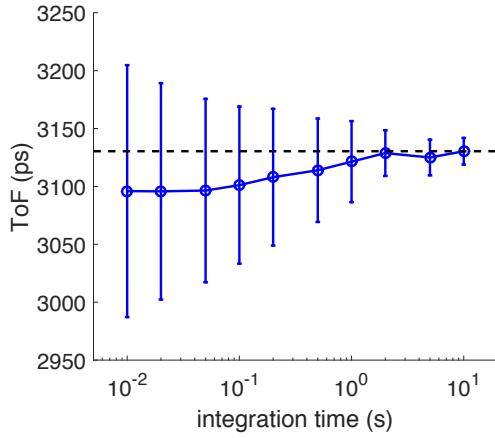
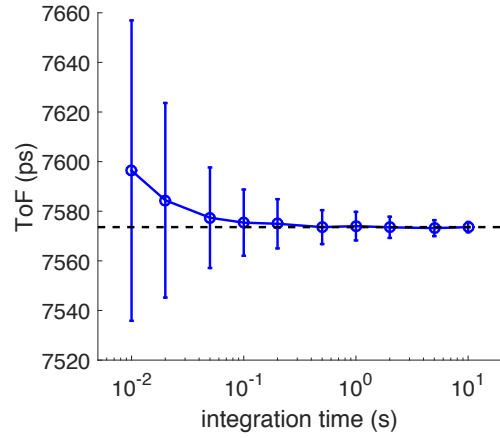


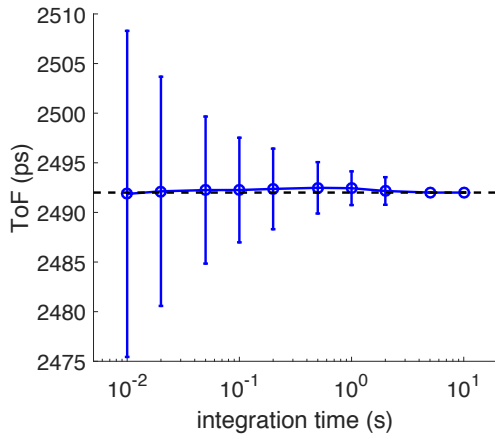
Figure 4.4: **First-returning photon estimation is stable even with short integration time.** The scene contains two planar objects, creating two main peaks in the transient response. Here we show the captured raw data in (a), and the processed data, estimated positions for the first-returning photons and the peak positions in the transient in (b).



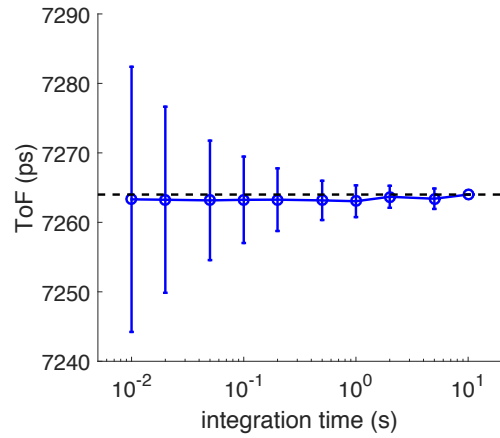
(a) peak position of first plane



(b) peak position of second plane



(c) first photon position of first plane



(d) first photon position of second plane

Figure 4.5: **Mean and standard deviation of estimated first-photon and peak positions.** We collected 10000 SPAD readout of 0.01s integration time. We combine different numbers of readout and estimate the first-returning photon and the peak positions of the two planes. We show the average and standard deviation of the estimation with respect to the integration time.

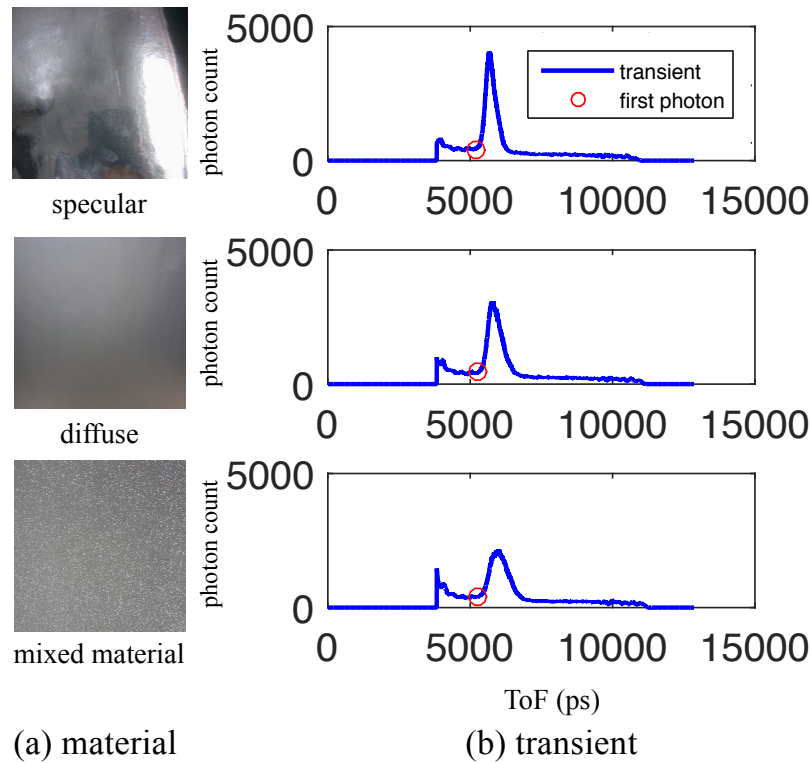


Figure 4.6: **First-returning photon estimation is invariant with respect to reflectance.** For fixed lighting and sensing positions, when imaging the same planar object with different material reflectance, the transient responses appear differently. However, the time-of-flight of the first-returning photon remains the same.

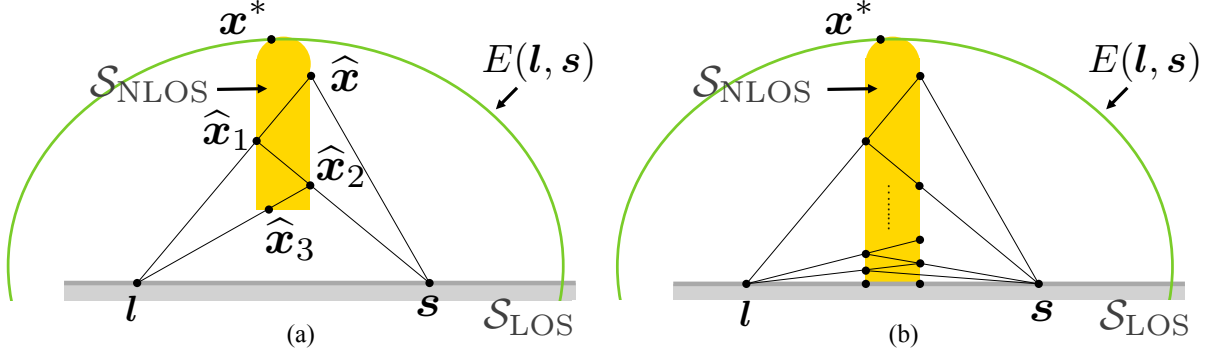


Figure 4.7: **Proof for Observation 1.** (a) There exists a visible point that creates a shorter light path, which leads to a violation of the definition of the first-returning photon. (b) Specialized case that needs to be excluded.

$$E(l, s) = \{x \mid \|x - l\| + \|x - s\| = d(l, s)\}.$$

Since the first-returning photon traverses the shortest path associated with l and s , this creates spatial constraint on occupancy of the NLOS object, which can be explained in the following observation.

Observation 1. There are no NLOS scene points in the interior of the ellipsoid $E(l, s)$.

Proof. Suppose that there exists a NLOS scene point \hat{x} inside the ellipsoid $E(l, s)$. If \hat{x} is visible to both l and s , then a three-bounce light path from l to \hat{x} to s will create a shorter light path, which contradicts the definition of the first-returning photon.

Suppose that \hat{x} is not visible to l . Then there exists an occluder \hat{x}_1 that blocks \hat{x} from l ; the occluder \hat{x}_1 lies on the line joining \hat{x} and l and is visible to l (see Figure 4.7(a)). If \hat{x}_1 is not visible to s , then we repeat the process to find a point \hat{x}_2 that is visible to s , and so on. We define the path length caused by the occluder \hat{x}_i to be

$$\|\hat{x}_i - l\| + \|\hat{x}_i - s\| = \tau(\hat{x}_i) = \tau_i.$$

We observe that $\tau_1 > \tau_2 > \tau_3 > \dots$, that is, the path length decreases each time instant. Given that $\{\tau_i\}$ is bounded from below by $\|l - s\|$, and is decreasing, it will converge via the monotone convergence theorem.

Let $\tau^* = \lim_{i \rightarrow \infty} \tau_i$ be the converged value. When $\tau^* > \|l - s\|$, it is easily shown that \hat{x}_i also converges and the converged point will necessarily be visible to both l and s (if not, we repeat the process above and the path length decreases). We can hence find a point visible to both l and s with a shorter light path than the first-returning photon, which is a contradiction.

If $\tau^* = \|l - s\|$, then \hat{x}_i lies on the line connecting l and s , and can potentially oscillate — this only happens in the scenario of Figure 4.7 (b) when l and s are occluded from each other. This scenario is avoided by assuming that the points are visible to each other — a scenario that is entirely consistent with our envisioned operating conditions. \square

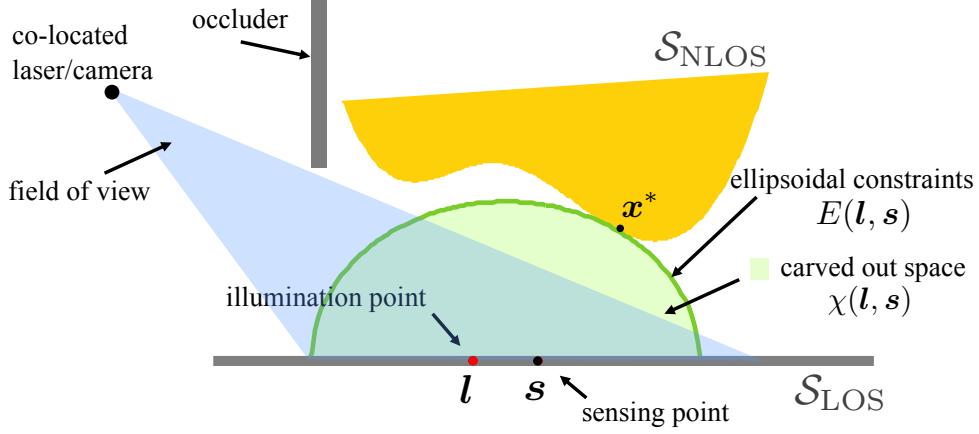


Figure 4.8: **No NLOS scene within the ellipsoid.** The ToF of the first-returning photon at each illumination and sensing pair creates an ellipsoidal constraint. Since we measure the shortest path between l and s , there should be no other NLOS scene with a shorter light path, and thus we carve out space inside the ellipsoid.

Defining the interior of $E(l, s)$ as

$$\chi(l, s) = \{x \mid \|x - l\| + \|x - s\| < d(l, s)\}, \quad (4.5)$$

we observe that

$$\chi(l, s) \cap \mathcal{S}_{\text{NLOS}} = \phi.$$

This observation is visualized in Figure 4.8.

Observation 1 constrains space where the NLOS scene $\mathcal{S}_{\text{NLOS}}$ can exist. We will build our observation on how to use local smoothness assumption to estimate the NLOS scene. First, we discuss the assumption that the shortest path is generated by a single NLOS point.

The uniqueness of the shortest path. Given l and s , the shortest three-bounce path between them is assumed to be unique, i.e., we assume that there exists only one NLOS scene point $x^*(l, s)$ such that

$$d(l, s) = \|x^*(l, s) - l\| + \|x^*(l, s) - s\|.$$

For simplicity, we will denote $x^*(l, s)$ simply as x^* .

Recall that given (4.1), the NLOS scene point(s) that contribute to the three-bounce light transient $I(t_0; l, s)$ lie on an ellipsoid; further, when t_0 is increased, the ellipsoid increases in size. Consider the point(s) of contact between the NLOS scene and the ellipsoid when t_0 is gradually increased; the shortest path between l and s is unique only when the contact is at a single location (see Figure 4.9). This is always the case when the NLOS scene is a convex shape.

There are instances of non-convex NLOS scenes that violate this uniqueness assumption. In fact, given l and s , it is not hard to construct a NLOS scene such that there are multiple points of contact to the ellipsoid. However, this is not a generic condition in that it requires very careful design of the NLOS scene for a given LOS scene pair; hence, if we perturb l and s , the symmetry of the scene is invariably broken and we recover uniqueness of the shortest path. In practice, this

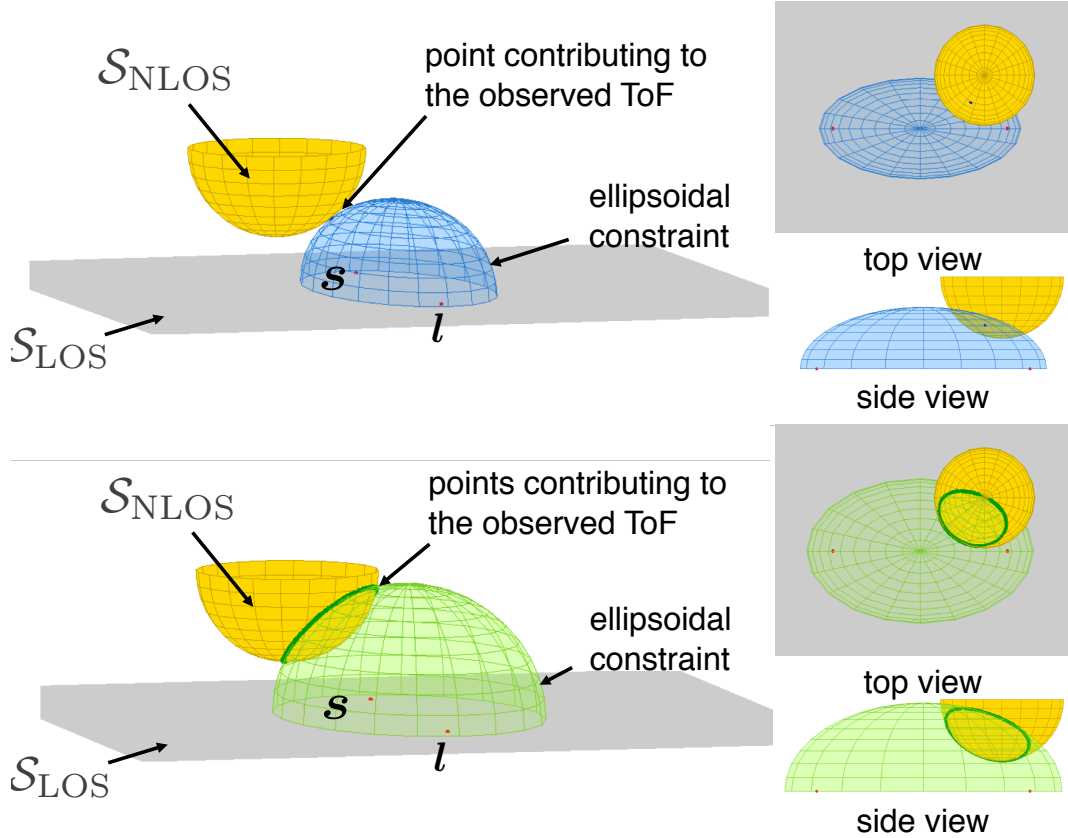


Figure 4.9: **A single scene point contributes to the first-returning photon.** The time-of-flight of the first-returning photon corresponds to the smallest ellipsoid that intersects with the NLOS scene. For a convex NLOS scene, the intersection will only contain one point.

implies that non-unique shortest path scenarios occur, at best, for a tiny subset of LOS scene pairs and can be handled as outliers. An example of non-unique shortest light paths is shown in Fig. 4.10(a). However, when we move the illumination and sensing points — even slightly — the shortest light path becomes unique (see Figure 4.10(b)).

By using additional local smoothness assumption regarding the local neighborhood of the unique scene point x^* , we find additional geometric constraints.

Observation 2. Suppose that NLOS scene is locally smooth at x^* . Then, the (unique) supporting plane at x^* is tangential to the ellipsoid $E(l, s)$.

Proof. Suppose that the supporting plane to the NLOS scene at x^* is not tangential to the ellipsoid $E(l, s)$. Then, given local smoothness, we can show that there are NLOS scene points in an infinitesimal neighborhood of x^* that belong to the interior of the ellipsoid, $\chi(l, s)$. This contradicts Observation 1. \square

We visualize this scenario in Figure 4.11. It is also worth noting that when local smoothness is violated — for example, at corners — the supporting plane to the NLOS object is not unique.

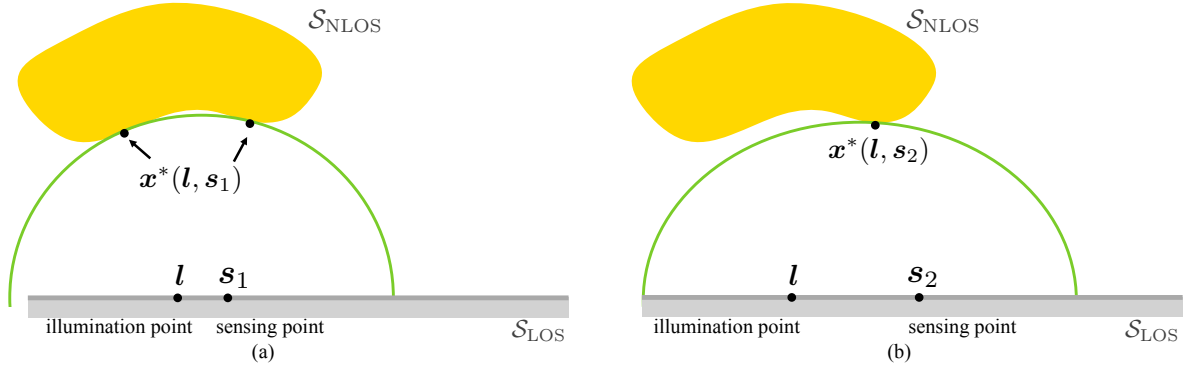


Figure 4.10: **The uniqueness of shortest three-bounce light path.** (a) We show an example where there are multiple scene points contributing to the shortest path. That is, the NLOS scene intersects with the ellipsoidal constraint $E(l, s_1)$ at multiple points. Even though this means that the shortest path between l and s_1 is not unique, when we move the illumination or the sensing point, this is no longer the case. (b) When we move the sensing to s_2 , now the NLOS scene only intersect with the ellipsoidal constraint at one location.

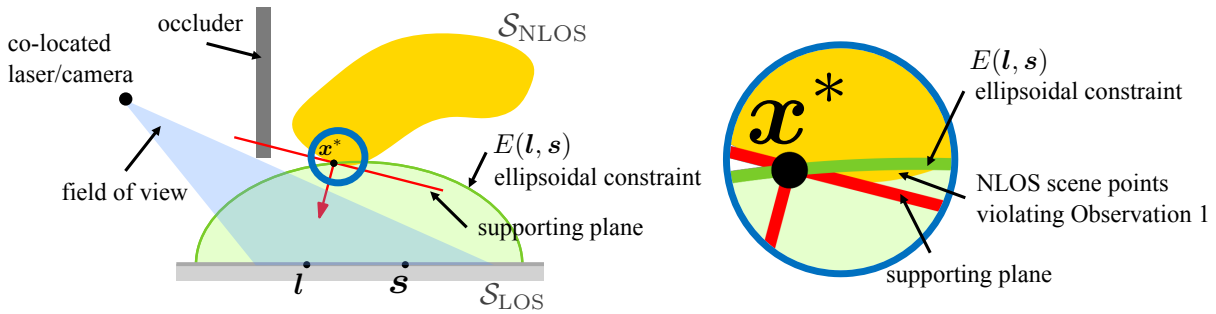


Figure 4.11: **When Observation 2 is violated, there exist NLOS scene points violating Observation 1.** If the supporting plane at x^* is not tangential to the ellipsoid $E(l, s)$, there will exist NLOS scene points belonging to the interior of the ellipsoid.

Observation 2 also implies that the supporting planes at x^* to both the NLOS scene and the ellipsoid are identical. This provides us with an explicit expression for the surface normal of the NLOS object at x^* .

Observation 3. Under the local smoothness assumption of the NLOS scene at x^* , the surface normal $\mathbf{n}(x^*)$ is the angular bisector of the vectors from x^* to the illumination spot l and sensing spot s , respectively; that is,

$$\mathbf{n}(x^*) \propto \frac{l - x^*}{\|l - x^*\|} + \frac{s - x^*}{\|s - x^*\|}. \quad (4.6)$$

Observation 3 is directly following the property of ellipsoids. Also, since the light path of interest corresponds to first-returning photon, it is the shortest light path, which follows Fermat's principle. Therefore, locally, the path will follow a mirror reflection.

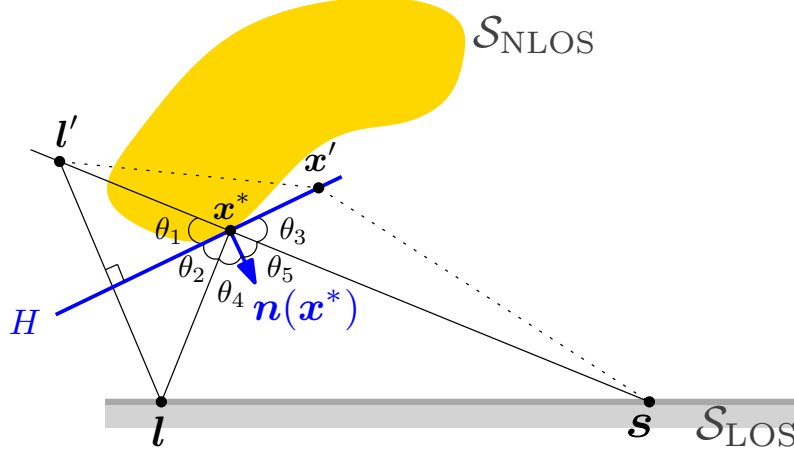


Figure 4.12: **Geometric interpretation of the normal being the angle bisector.** l' is the mirror image of l w.r.t H . Then, $\theta_1 = \theta_2$ and the distance from any point on H to l is the same as the distance to l' . Since the shortest path from l' to s is the straight line connecting them, x^* lies on this straight line; otherwise, there exists an alternate path that is shorter. Therefore, $\theta_3 = \theta_1$. We can conclude that $\theta_4 = \pi/2 - \theta_2 = \pi/2 - \theta_1 = \pi/2 - \theta_3 = \theta_5$. The normal $\mathbf{n}(x^*)$ is the angular bisector.

The following geometric interpretation of Observation 3 is useful for the shape recovery algorithms in Section 4.5. Given the supporting plane H at x^* , we find the mirror image of the illumination point l with respect to H ; we denote the mirror image as l' . When the surface normal \mathbf{n} satisfies (4.6), then it can be shown that the straight line joining the mirror image l' to s passes through x^* . A brief proof of this is presented in Figure 4.12.

In the following, we provide the algebraic proof of Observation 3.

Proof. Recall that from the ellipsoid property, the surface normal of supporting plane of an ellipsoid is the angular bisector. We now prove that the normal at x^* is also the angle bisector and hence, the supporting planes to the NLOS surface and the ellipsoid at x^* are one and the same.

Consider the setup in Figure 4.13. Light emitted from a source at l bounces off an object and reaches the point s . The shortest among all such paths passes through the point x^* on the object. Given this, we prove that the surface normal at x^* — assuming that the object is locally smooth — satisfies

$$\mathbf{n} \propto \frac{\mathbf{l} - \mathbf{x}^*}{\|\mathbf{l} - \mathbf{x}^*\|} + \frac{\mathbf{s} - \mathbf{x}^*}{\|\mathbf{s} - \mathbf{x}^*\|},$$

or equivalently, the surface normal at x^* is the angular bisector of the vectors to l and s , from x^* .

Since the NLOS object is locally smooth, therefore, points in an infinitesimal local neighborhood of the discovered x^* can be represented as

$$\mathbf{x}' = \mathbf{x}^* + \alpha \mathbf{n}_1 + \beta \mathbf{n}_2, \quad (4.7)$$

where \mathbf{n}_1 and \mathbf{n}_2 are the basis spanning the local planar approximation. That is, $\langle \mathbf{n}, \mathbf{n}_1 \rangle = 0$, $\langle \mathbf{n}, \mathbf{n}_2 \rangle = 0$, and $\langle \mathbf{n}_1, \mathbf{n}_2 \rangle = 0$, where \mathbf{n} is the surface normal. We can now compute the distance

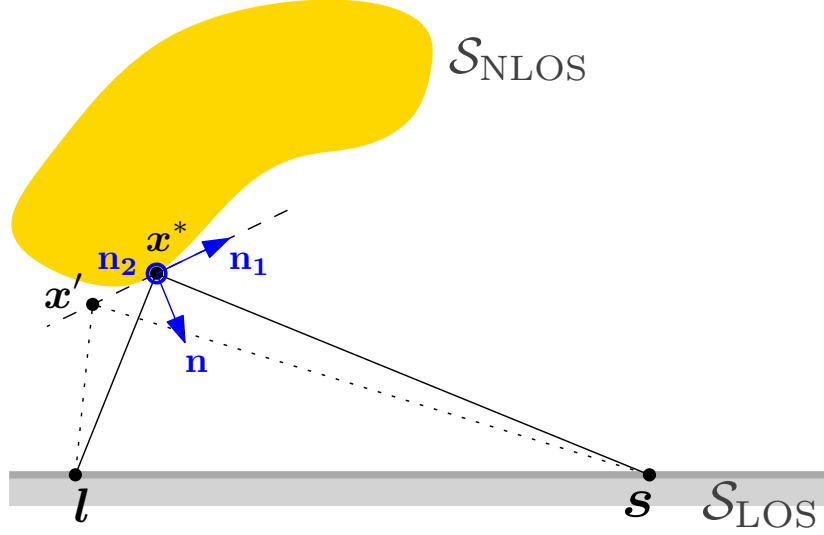


Figure 4.13: **Normal estimation of the recovered NLOS scene point.** The normal of the scene patch corresponding to the first photon is the angle bisector. For all other points \mathbf{x}' in the neighboring area, the time of flight will be larger compared to \mathbf{x}^* .

from l to s via points on the local planar patch at \mathbf{x}' .

$$\begin{aligned}\tau(\mathbf{x}') &= \|\mathbf{x}' - l\| + \|\mathbf{x}' - s\| \\ &= \|\mathbf{x}^* + \alpha \mathbf{n}_1 + \beta \mathbf{n}_2 - l\| + \|\mathbf{x}^* + \alpha \mathbf{n}_1 + \beta \mathbf{n}_2 - s\|\end{aligned}$$

Since the shortest path in this local neighborhood happens for $\mathbf{x}' = \mathbf{x}^*$, the derivative of $\tau(\cdot)$ with respect to α and β is zero at $\mathbf{x}^* = \mathbf{x}'$, or equivalently, when $\alpha = \beta = 0$.

$$\begin{aligned}\frac{\partial \tau(\mathbf{x}')}{\partial \alpha} &= \frac{\mathbf{n}_1^T (\mathbf{x}^* + \alpha \mathbf{n}_1 + \beta \mathbf{n}_2 - l)}{\|\mathbf{x}^* + \alpha \mathbf{n}_1 + \beta \mathbf{n}_2 - l\|} + \frac{\mathbf{n}_1^T (\mathbf{x}^* + \alpha \mathbf{n}_1 + \beta \mathbf{n}_2 - s)}{\|\mathbf{x}^* + \alpha \mathbf{n}_1 + \beta \mathbf{n}_2 - s\|} \\ \frac{\partial \tau(\mathbf{x}')}{\partial \beta} &= \frac{\mathbf{n}_2^T (\mathbf{x}^* + \alpha \mathbf{n}_1 + \beta \mathbf{n}_2 - l)}{\|\mathbf{x}^* + \alpha \mathbf{n}_1 + \beta \mathbf{n}_2 - l\|} + \frac{\mathbf{n}_2^T (\mathbf{x}^* + \alpha \mathbf{n}_1 + \beta \mathbf{n}_2 - s)}{\|\mathbf{x}^* + \alpha \mathbf{n}_1 + \beta \mathbf{n}_2 - s\|}\end{aligned}$$

The minimum is reached when $\mathbf{x}' = \mathbf{x}^*$. That is, the derivative reaches 0 when $\alpha = 0$ and $\beta = 0$.

$$\begin{aligned}\frac{\partial \tau(\mathbf{x}')}{\partial \alpha} \Big|_{\alpha=0, \beta=0} &= \mathbf{n}_1^T \left(\frac{\mathbf{x}^* - l}{\|\mathbf{x}^* - l\|} + \frac{\mathbf{x}^* - s}{\|\mathbf{x}^* - s\|} \right) = 0 \\ \frac{\partial \tau(\mathbf{x}')}{\partial \beta} \Big|_{\alpha=0, \beta=0} &= \mathbf{n}_2^T \left(\frac{\mathbf{x}^* - l}{\|\mathbf{x}^* - l\|} + \frac{\mathbf{x}^* - s}{\|\mathbf{x}^* - s\|} \right) = 0\end{aligned}$$

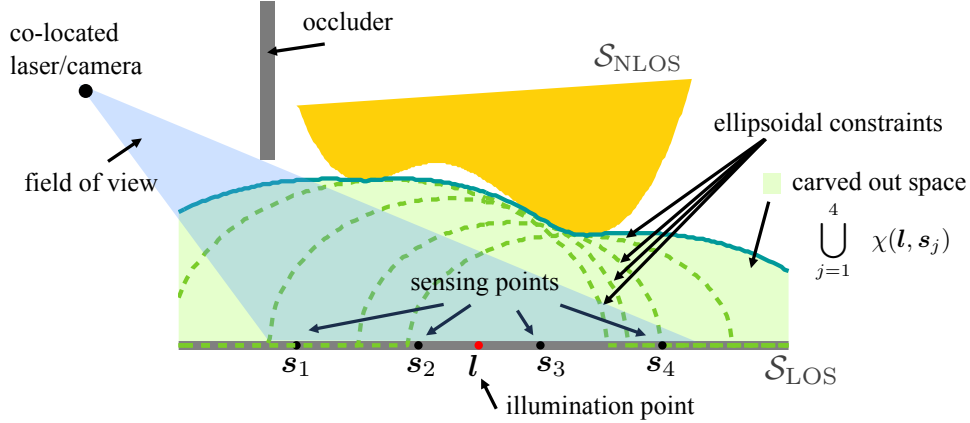


Figure 4.14: **Space carving for NLOS imaging.** From Observation 1, we carve out space inside each ellipsoid. The space carved region will contain no NLOS scene points.

The relationships above suggest that the vector $\left(\frac{\mathbf{x}^* - \mathbf{l}}{\|\mathbf{x}^* - \mathbf{l}\|} + \frac{\mathbf{x}^* - \mathbf{s}}{\|\mathbf{x}^* - \mathbf{s}\|} \right)$ is perpendicular to both \mathbf{n}_1 and \mathbf{n}_2 and hence, it must be aligned along the surface normal since the surface normal is perpendicular to both \mathbf{n}_1 and \mathbf{n}_2 . Therefore,

$$\mathbf{n} \propto \frac{\mathbf{x}^* - \mathbf{l}}{\|\mathbf{x}^* - \mathbf{l}\|} + \frac{\mathbf{x}^* - \mathbf{s}}{\|\mathbf{x}^* - \mathbf{s}\|} \quad (4.8)$$

Here we show that the normal at \mathbf{x}^* is equal to the angle bisector, which is also the normal of the supporting plane of the ellipsoid at \mathbf{x}^* . Therefore, the supporting plane is tangential to the ellipsoid. □

4.5 Shape from first-returning photons

We propose three algorithms utilizing the finding in the three observations in Section 4.4 — the first for carving out the space the NLOS object cannot occupy and the latter for recovering the location and surface normals of NLOS scene points that generate the first-returning photons.

Algorithm 3 (Space carving for NLOS imaging).

We extend Observation 1 by incorporating space carving constraints from all pairs of illumination and sensor points. Specifically, the NLOS scene $\mathcal{S}_{\text{NLOS}}$ cannot lie within the union of the individual ellipsoids, i.e.,

$$\bigcup_{\mathbf{l}, \mathbf{s} \in \mathcal{S}_{\text{LOS}}} \chi(\mathbf{l}, \mathbf{s}) \cap \mathcal{S}_{\text{NLOS}} = \phi,$$

where $\chi(\mathbf{l}, \mathbf{s})$ is defined in (4.5). We illustrate space carving from multiple first-returning photons in Figure 4.14, when illuminating a single LOS spot.

Algorithm 4 (NLOS shape recovery under the local planarity assumption).

We assume that a small neighborhood of LOS sensing spots $\{\mathbf{s} \in \Omega\}$ receive first-returning photons from a locally-planar NLOS scene patch when illuminating the LOS wall at \mathbf{l} . Therefore,

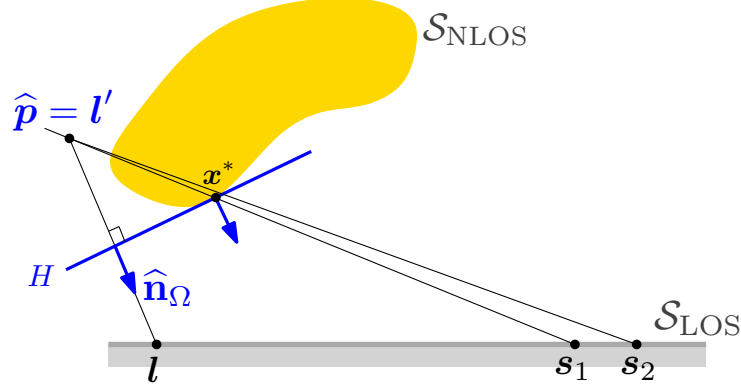


Figure 4.15: **NLOS shape and normal under the local planarity assumption.** If we assume the first-returning photons from neighboring sensors are from the same local planar patch, then they share the same mirrored light source location l' . Thus, finding l' can subsequently lead to the recovery of NLOS scene point and surface normal.

we can combine the measured ToFs to infer the location and orientation of the locally-planar patch. We achieve this by estimating the location of the mirror image of the illumination point (see Figure 4.15).

Recall that the length of the shortest path from l to s is equal to the distance between the mirror image l' and s , $\|l' - s\|$; hence, $d(l, s) = \|l' - s\|$. Given the collection $\{d(l, s) \mid s \in \Omega\}$, we can solve for the location of the mirrored image as

$$\min_{\mathbf{p}} \sum_{s \in \Omega} (d(l, s) - \|\mathbf{p} - s\|)^2. \quad (4.9)$$

The optimization problem is non-convex; we solve it using gradient descent techniques and, thus, the result depends heavily on the initialization. We initialize with the algebraic minimizer of the objective function. Without loss of generality, we assume the sensors are all on the same plane $z = 0$. That is, $s_j = (s_j^x, s_j^y, 0)$. We seek to find the location for the mirrored light source $l' = (l^x, l^y, l^z)$.

By squaring the path length, we get

$$d^2(l, s_j) = (l^x - s_j^x)^2 + (l^y - s_j^y)^2 + (l^z - 0)^2 \quad (4.10)$$

For simplicity, we will denote $d(l, s_j)$ as d_j .

We use another sensor location s_i that belongs to the same local neighborhood to jointly estimate (l^x, l^y, l^z) .

$$d_i^2 = (l^x - s_i^x)^2 + (l^y - s_i^y)^2 + (l^z - 0)^2 \quad (4.11)$$

(4.11) - (4.10), we get

$$\begin{aligned}
d_i^2 - d_j^2 &= \left[(l^x - s_i^x)^2 - (l^x - s_j^x)^2 \right] + \left[(l^y - s_i^y)^2 - (l^y - s_j^y)^2 \right] \\
&= (2l^x - s_i^x - s_j^x)(s_j^x - s_i^x) + (2l^y - s_i^y - s_j^y)(s_j^y - s_i^y) \\
&= 2(s_j^x - s_i^x)l^x + 2(s_j^y - s_i^y)l^y - \left[(s_j^x)^2 - (s_i^x)^2 \right] - \left[(s_j^y)^2 - (s_i^y)^2 \right]
\end{aligned}$$

We can see that the relationship is linear in l^x and l^y . Therefore, we can construct a linear system. Here, for ease of interpretation, we use $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m$ to represent sensors in the local neighborhood.

$$2 \begin{bmatrix} (s_1^x - s_2^x) & (s_1^y - s_2^y) \\ (s_1^x - s_3^x) & (s_1^y - s_3^y) \\ \vdots & \vdots \\ (s_1^x - s_m^x) & (s_1^y - s_m^y) \end{bmatrix} \begin{bmatrix} l^x \\ l^y \end{bmatrix} = \begin{bmatrix} d_2^2 - d_1^2 + [(s_1^x)^2 - (s_2^x)^2] + [(s_1^y)^2 - (s_2^y)^2] \\ d_3^2 - d_1^2 + [(s_1^x)^2 - (s_3^x)^2] + [(s_1^y)^2 - (s_3^y)^2] \\ \vdots \\ d_m^2 - d_1^2 + [(s_1^x)^2 - (s_m^x)^2] + [(s_1^y)^2 - (s_m^y)^2] \end{bmatrix}$$

(l^x, l^y) can be solved using pseudo inverse. Finally, the mirrored light source location can be solved

$$l^z = \sqrt{d_j^2 - (l^x - s_j^x)^2 - (l^y - s_j^y)^2}$$

Once we have an estimate for the mirror image $\hat{\mathbf{p}}$, the estimate of the surface normal to the planar patch is given as

$$\hat{\mathbf{n}}_\Omega = \frac{(\mathbf{l} - \hat{\mathbf{p}})}{\|\mathbf{l} - \hat{\mathbf{p}}\|}.$$

We can also identify $(\mathbf{l} + \hat{\mathbf{p}})/2$ as a point on the supporting plane H , which gives us the equation of the planar patch. We now identify points on the plane that produce the first-returning photons by intersecting this plane with the straight lines from $\hat{\mathbf{p}}$ to each of the points $\mathbf{s} \in \Omega$.

An assumption of Algorithm 4 is that local sensing points receive first-returning photons from the same local patch. However, as shown in Figure 4.16(b), when there are multiple objects in the scene, neighboring sensor locations receive first-returning photons from different sources. This causes an error in the shape recovery. The solution is to distinguish the sources that generate first-returning photon. Separating source of first-returning photons is possible since the ToF and the sensor locations have a quadratic-like structure (Figure 4.16(b)). Based on this observation, we find the ToF model of first-returning photons and propose Algorithm 5.

Algorithm 5 (NLOS shape recovery using the union of quadratic models).

The assumption of local planarity uses neighboring measurements to jointly estimate the location of the mirrored illumination. The results will be affected by the choice of neighborhood size; a small neighborhood size has a noisy estimate, whereas a large neighborhood size provides a more accurate estimate but may suffer from model mismatch. It would be most beneficial to automatically cluster as many measurements as possible while making sure the model is still the same.

We make the following two observations about the ToF measurements generated by two models:

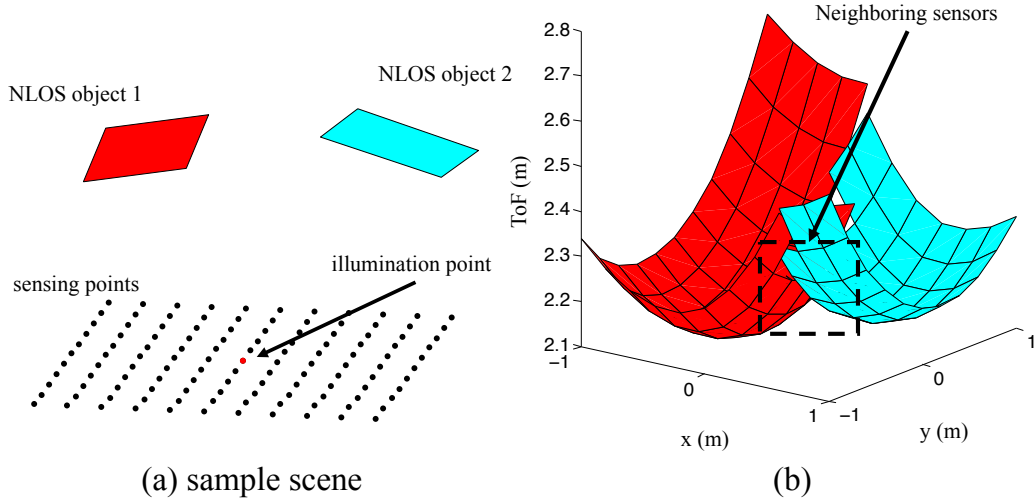


Figure 4.16: **Failure case for Algorithm 4.** (a) A sample scene to illustrate when Algorithm 4 will fail. The scene consists of two planar patches. In (b) x and y axes correspond to the sensor locations and the z axis corresponds to the ToF of the first-returning photon with color indicating the source of the first-returning photon. There exist some local neighboring sensing locations that receive first-returning photons from different planar patches causing assumption of Alg 4 to fail.

- *Plane.* If the first-returning photons arriving at different sensing points are generated from the same plane, the unknown to be estimated will be the mirrored position of the illumination point $\mathbf{p}_{\text{plane}} = [x_{\text{plane}}, y_{\text{plane}}, z_{\text{plane}}]$ (see Figure 4.17 (a)).

$$d(\mathbf{l}, \mathbf{s}) = \|\mathbf{p}_{\text{plane}} - \mathbf{s}\|$$

$$d^2(\mathbf{l}, \mathbf{s}) = (x_{\text{plane}} - s^{(x)})^2 + (y_{\text{plane}} - s^{(y)})^2 + (z_{\text{plane}} - s^{(z)})^2.$$

The unknowns, x , y , and z , are quadratic with respect to the measured time-of-flight and the location of the sensing spot.

- *Point.* If the first-returning photons arriving at different points are generated from the same point, the unknown to be estimated will be the point itself.

$$d(\mathbf{l}, \mathbf{s}) = \|\mathbf{p}_{\text{point}} - \mathbf{l}\| + \|\mathbf{p}_{\text{point}} - \mathbf{s}\|$$

Without loss of generality, we can shift the whole system so that \mathbf{l} is at the origin. By introducing an additional variable, $R = \|\mathbf{p}_{\text{point}}\|$:

$$[d(\mathbf{l}, \mathbf{s}) - R]^2 = (x_{\text{point}} - s^{(x)})^2 + (y_{\text{point}} - s^{(y)})^2 + (z_{\text{point}} - s^{(z)})^2.$$

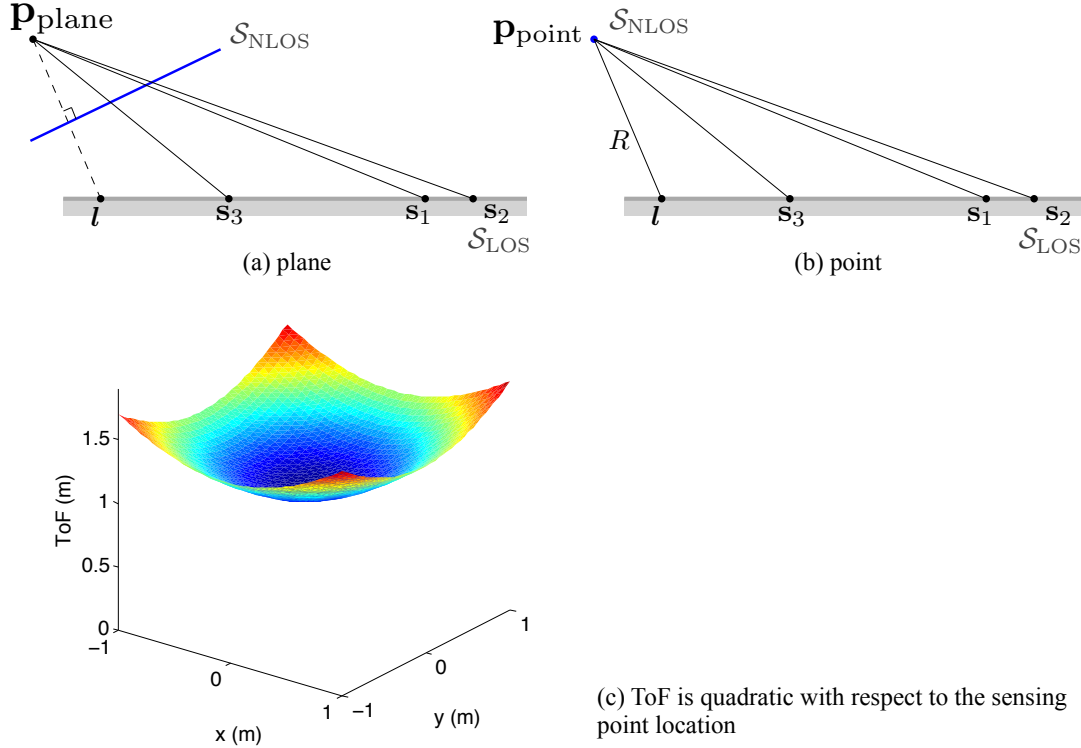


Figure 4.17: **Quadratic models for NLOS imaging.** (a) If the NLOS object is a plane, the ToF of the first-returning photon coming from this plane is the distance from the mirrored illumination to the sensing point. (b) If the NLOS object is a point, the ToF of the first-returning photon coming from this point is the distance from the point to the sensing position plus a constant R . (c) We plot the x, y coordinate of the sensing point and the ToF of the received first-returning photon in z coordinate.

The unknown, x_{point} , y_{point} , z_{point} , and R , are quadratic with respect to the known quantities. Here we treat R as a variable; however, it is related to $\mathbf{p}_{\text{point}}$ (see Figure 4.17 (b)).

For the two sources, we can use 4 measurements to form an initial estimate. We first randomly select a point, then randomly select 3 points in proximity to it since for most cases, neighboring sensing positions receive first-returning photons from the same source. We use these 4 points to form an initial estimate and find other inliers to jointly form a model candidate for RANSAC. We use the average error between the measured ToF and the ToF from the estimated models to determine the best-fitted model. By using RANSAC we can find the measurements of different sensing points to be used together. This can also help to handle outliers since we can reject samples based on how well the model explains the data. Between the two models (plane and point), we choose the one that generates a smaller average distance between the ToF of first-returning photons and the ToF of the recovered points.

In a real scene, there will exist more than planes and points. For any smooth surface, we approximate a local curved patch as a local planar patch.

4.6 Experiments

We demonstrate the effectiveness of our proposed algorithms through both synthetic data and realistic setup. We show how space carving can help reduce the free space where the NLOS object resides. Also, we use local planarity to localize scene points and estimate the surface normal. Finally, we verify our method based on RANSAC to fit local planar patches and corner points that generate first-returning photons. Unless otherwise noted, the NLOS scene is Lambertian.

4.6.1 Simulation results

Rendering setup. We used the code from [58] to render the light transient. Rendering three-bounce light paths is a time-consuming process, therefore, to improve the efficiency of rendering, we modify the imaging setup by placing an omni-directional light source and an omni-directional sensor on the LOS scene and rendering single-bounce light paths. This setup is equivalent to NLOS imaging.

Space carving for NLOS imaging. We demonstrate NLOS space carving for the scene, shown in Figure 4.18, consisting of a bunny in a room occluded from the laser/camera location. We randomly select 100 pairs of illumination and sensing points that are visible to the laser and camera center. We also make sure that the illumination and sensing points are visible to each other to avoid scenario described in Figure 4.7(b).

The result of the space carving is visualized in Figure 4.19 with two different number of measurement used. We observe that the bunny is outside the carved region. Further, the volume of the entire NLOS scene is 5.56 cubic meters. With space carving, we identify free space in the NLOS scene and hence, the resulting NLOS scene volume with 100 pairs of illumination and sensing points is reduced to 0.44 cubic meters, which is only 7.86% of the original space. The volume that the bunny occupies is 0.20 cubic meters. Our method successfully decreases the space of possible NLOS scene. However, due to the complex shape of the bunny, some space will not be carved out.

Space recovery with different reflectances. The ToF is solely a function of the geometry of the NLOS scene. Thus, our proposed method can work for NLOS scene with unknown reflectance as oppose to prior works that rely on the assumption of the NLOS scene being Lambertian. As shown in Figure 4.6, the ToF of first-returning photons of an object with different reflectance remains the same.

We further verify this claim by rendering the light transient of the NLOS scene with different reflectances, from purely Lambertian to highly specular reflectance. We control the reflectance via the parameters of the Blinn-Phong model. The scene setup is shown in Figure 4.20. The NLOS object is a sphere and we illuminate one point and sense at 957 points. As shown in Figure 4.21, the rendered ToF and hence, the recovered points, are largely invariant to the reflectances of the NLOS object.

Shape recovery with different noise levels. The scene contains a sphere where the location of the center \mathbf{o} and the radius r are known. We compute the recovery error by finding the average distance between the recovered point to the surface of the sphere.

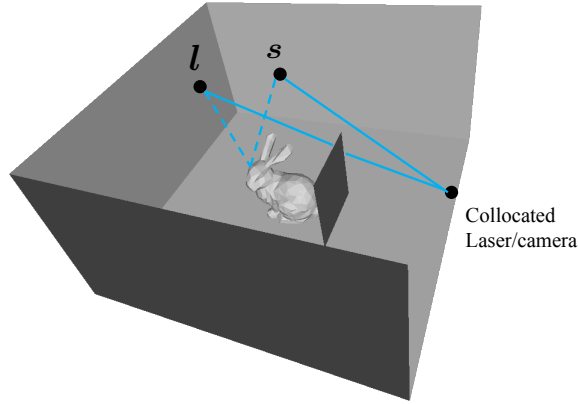


Figure 4.18: **Scene setup for demonstrating the space carving algorithm.** We place a bunny behind an occluder and orient the laser and SPAD toward different points on all the walls to image the NLOS bunny. For ease of visualization, we do not show the top wall.

$$E_x = \frac{1}{m} \sum_{i=1}^m | \|x_i - \mathbf{o}\| - r |,$$

where x_i is the i -th recovered point. As for the normal recovery error, we find the average angular error between the recovered normal and the normal of the projected point on the sphere.

$$E_{\mathbf{n}} = \frac{1}{m} \sum_{i=1}^m \cos^{-1} \left(\mathbf{n}_i^\top \frac{x_i - \mathbf{o}}{\|x_i - \mathbf{o}\|} \right),$$

where \mathbf{n}_i corresponds to the normal estimation of recovered point x_i .

As shown in Figure 4.22(a), we observe that the reconstruction error is roughly linear to the noise level. We show the recovered shapes in Figure 4.22(b) for visualization. We also show recovered normals in Figure 4.22(c) with color-coded normal.

Shape recovery with different neighborhood sizes. In (4.9), we use a local neighborhood to solve the mirrored light source location. By choosing a larger area, we include more measurements, thus the effect of noise can be alleviated. However, for very large neighborhoods, the locally planar assumption can be violated, causing large model misfit errors. We use the setup in Figure 4.20 and compare the reconstruction results with respect to different neighborhood sizes. We observe in Figure 4.22 that the error first reduces with increasing neighborhood size and subsequently, increases due to the violation of the local planarity assumption.

Shape recovery with the union of quadratic models. We can use Algorithm 5 to fit quadratic models to the time-of-flight measurements of the first-returning photon. We repeat this process until there is no more measurement that fits any models. The advantage of using Algorithm 5 is that we do not need to tune the neighborhood size parameter as required in Algorithm 4. However, for a scene that violates the assumption that first-returning photons are generated by local planes

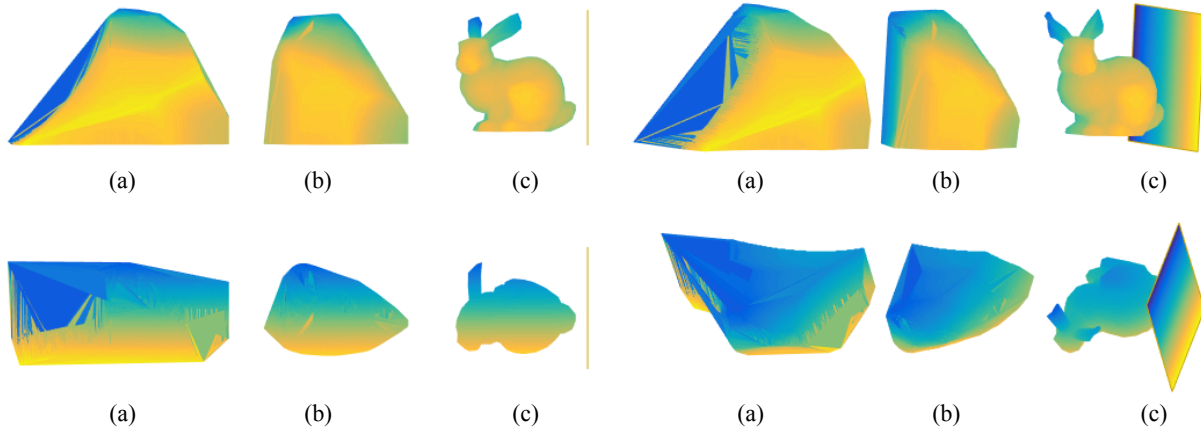


Figure 4.19: **Space carving algorithm results.** We present the space carving results alongside with the ground truth in four views. (a) We show the space carving result using 25 pairs of illumination and sensing spots. (b) Space carving result using 100 pairs of illumination and sensing spots. (c) Ground truth.

or a single point, we can see that the Algorithm can only pick up even fewer parts of the scene (see Figure 4.22 [iii,iv]). Still, Algorithm 5 has a great advantage when the scene is truly composed of multiple distinct planar objects. We demonstrate this with real-world examples.

4.6.2 Real scene

In the following experiments, we used four data sources that use a pulsed laser and a SPAD to gather the light transient of a NLOS scene:

- *Dataset 1.* Buttafava *et al.* [15] collect three-bounce light paths using 185 illumination points and sensing at a single spot on the wall (see Figure 4.23). There are three NLOS objects, a T shape, a larger square behind the T shape, and a smaller square. There is an additional surface due to the optical setup.
- *Dataset 2.* Pediredla *et al.* [102] measure the NLOS scene by placing planes at different locations and orientations. They focus the SPAD to a single location and use a galvo mirror to direct the laser to 48 locations.
- *Dataset 3.* We use the setup described in [102] and collect the NLOS scene with more varieties. We focus the SPAD at a single location and illuminate the scene at 168 lighting locations. In this dataset, we explore different aspects of NLOS imaging. This includes changing the integration time, planar NLOS objects, curved NLOS objects, and multiple NLOS objects.
- *Dataset 4.* We collect this dataset to explore the trade-off between integration time and tracking. The acquisition setup is described in [102]. The SPAD is focused at a single location and we illuminate 20 points on the wall with a laser. The NLOS object is a plane mounted on a mechanical translation and rotation stage. We collect data with three kinds of movement, 2 translations and 1 rotation.

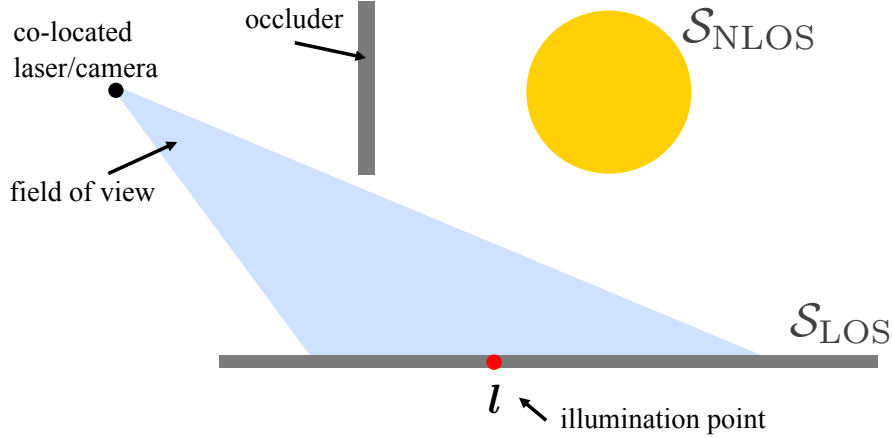


Figure 4.20: **Shape recovery scene setup.** Here we show the side view of the setup. The NLOS scene consists of a sphere. We image the sphere through 1 illumination point and 957 sensing points.

We collect nine integration times, which are 1, 2, 5, 10, 20, 50, 100, 200, 500 ms. We raster scan the 20 illumination points, each with the desired integration time. And we repeat this process for 30 seconds for translation movements and 60 seconds for rotation movement. During the acquisition, the NLOS object is constantly moving, controlled by the mechanical stage.

From the 5D light transient, we find the ToF of the first-returning photon by finding the first time instant that photon count increases significantly. In all these datasets, a single-pixel SPAD is used, thus the sensing spot is fixed. With a galvo mirror, laser light is directed to different parts of the wall, creating different illumination points. Because of Helmholtz reciprocity, we can switch the role of illumination and sensing, then apply the shape recovery algorithm. That is, instead of finding the mirrored illumination point, we find the mirrored sensing point.

Shape recovery with different integration time. We verify our claim that methods based on first-returning photon is robust to short acquisition time. In Figure 4.24, we show the result of three different integration time for each illumination spot. We can see that the estimated first-returning photons are invariant with integration time. As expected, the ToF of extremely short integration time, 1 ms, is noisier. Along with a robust algorithm (Algorithm 5), we can still recover an accurate point cloud estimation of the hidden object.

Shape recovery with planes. We quantify the recovered results by computing error in position and normal estimation. We can compute the theoretical source of scene points that contribute to first-return photons from the ground truth scene description. The scene labeled with * means that the points are truly coming from a plane. For all other scenes, the scene points contributing to first-returning photon corresponds to edges of the scene. We test our ability to recover points on planes using two datasets, as shown in Table 4.2 and Table 4.3. We can see that even if we do not explicitly model edges in our algorithm, the algorithm does not completely fail. The recovered points will still lie closely to the plane; however, as is to be expected, the normal estimation will be more inaccurate.

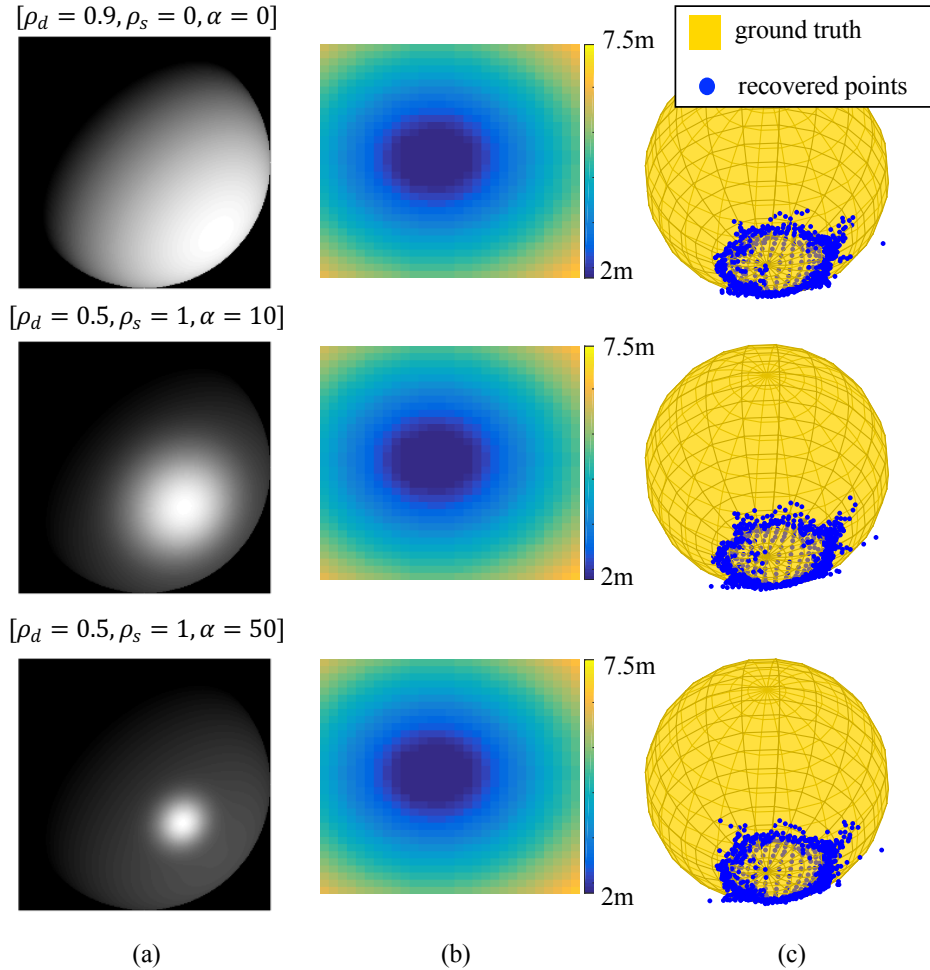


Figure 4.21: **Shape reconstruction under varying NLOS scene reflectance.** Here we show the results of 3 different material reflectances. (a) Reflectance visualization, shown by illuminating the sphere with a point light source. (b) The rendered ToF at each sensing location. (c) The recovered NLOS scene points.

Shape recovery with curved objects. We collect four scenes with the setup similar to Figure 4.20. Scene 1-3 are a yoga ball from close to the wall to far. Scene 4 is a small ball. As shown in Table 4.4, our recovered results can indicate well where the NLOS object is. We notice that the local planarity assumption fails in scene 4 because the radius of the ball is comparably smaller. We can see the normal estimation fail much worse in this case.

Shape recovery with multiple objects. Buttafava *et al.* [15] collect a setup containing multiple planar objects in the NLOS scene. This setup perfectly fits our model assumption of local planarity. We use a neighborhood size of 5 for shape recovery (Alg. 4) and filter out points based on the space carving constraint (Alg. 3). The recovered result is shown in Figure 4.25(a). Notice that we can only recover the two parts closest to the wall since the square behind the T shape cannot create first-returning photon observations.

By observing the 5D light transient (see Figure 4.23(b)), we can clearly see multiple peaks in the light transient. Those are first-returning photons generated by other surfaces that are farther from the wall, thus they will not be picked out by finding the first time instant when the photon count exceeds the threshold. Therefore, we use a signal processing method to find the peaks in the light transient then assign labels by clustering measurements of similar distance. As seen in Figure 4.25(b), we are able to recover more scene points.

As we can see in the color label auto-generated in Figure 4.25(b), there are instances where first-returning photons from different scene patches are clustered together incorrectly. From Alg. 5, we know that the structure of ToF coming from the same source will be quadratic, therefore, we can use RANSAC to pick up measurements that are from the same source automatically. This avoids the process of label assignment. As seen in Figure 4.25(c), we can recover scene points and identify different scene better. By considering the structure of the measurement, we adaptively cluster measurements and produce better results.

Notice that we recover each scene point independently. The recovered scene points lie on different planar scene objects, which are consistent to the setup described in [15]. Also, the recovered normals (see Figure 4.25) are roughly perpendicular to the point collection, which means our normal estimation is meaningful.

Shape recovery with multiple objects with ground truth scene description. To help evaluating the results of multiple algorithms, we collect a dataset that contains scenes with multiple objects along with the scene ground truth description. To evaluate the recovered results, we show the average distance to ground truth scene and the average normal estimation error. Since there are multiple objects in the scene, we use the closest ground truth object as ground truth for each recovered point.

As shown in Table 4.5, for most cases Alg 4 and 5 behave similarly. In scene 3 and 4, the major difference happens when hidden objects are closely placed. Labeling based on clustering will fail in this case and there will be some artifact between recovered points.

Tracking NLOS objects. We create a dataset that contains 3 different object movements: linear movement parallel to the wall, linear movement perpendicular to the wall, and rotating the NLOS object. The object is mounted on a mechanical stage that continuously moves during the acquisition time.

We fix the sensing position and raster scan the wall at 20 illumination points. We use Alg 5 to reconstruct parameter of a plane from the first-returning photon of the 20 positions. We show in Figure 4.26 the result of tracking the NLOS object. For visualization, we only show the intercept and the angle between the wall. We can qualitatively see that the recovered position of the plane is consistent with the movement of the stage. By plotting the estimation with respect to time, we see the frequency of the recovered motion is consistent across different acquisition times. However, as shown in forth row in Figure 4.26, for large acquisition time, since the object moves a lot during the data acquisition, there is no consistent estimation of the NLOS object.

4.7 Discussion

In this chapter, we show the acquisition advantages and geometric properties of first-returning photons in NLOS imaging. We provide both theoretical and experimental justification regarding the statistical behavior of first-returning photon estimates. We also identify novel constraints that arise from the study of shortest paths and design shape recovery algorithms that are simple and lightweight. To our knowledge, ours is the first technique that directly resolves NLOS shape and normals without solving complex inverse problems. We also create a dataset along with the ground truth scene description, which we believe will facilitate research in this field. In the following, we discuss several aspects of our proposed methods:

The sampling of the NLOS scene points. A drawback of using first-returning photon is that the NLOS scene points that produce the first-returning photon is a function of the scene geometry, i.e., we have limited control over the points that end up generating the first-returning photons. As a consequence, it is entirely possible that some scene points will never create first-returning photons. One such example is the surface on the top right corner of Figure 4.25. Only if we use signal processing techniques to extract more first-returning photons from the light transient can we recover the shape.

Failure of local planarity. The local planarity assumption is violated, often significantly, at depth discontinuities — for example, at edges and corners. Here, Alg. 5 is a good way to deal with such case that planar patches and corners can be identified and estimated separately. Since we can model the ToF from these two sources easily, using a RANSAC process to iteratively pick up the best-fitting model can help grouping correct subsets of measurements.

NLOS reflectances, smoothness, and convexity. From the Observation 3, the light path of the first-returning photon follows the mirror direction. This means that specular BRDF is actually most advantageous for smooth and convex objects, since all photons will be first-returning photons. Lambertian reflectance, on the contrary, creates light paths that belong to the tail of the light transient. This makes determining the ToF of the first-returning photon harder.

Specular BRDF is not always favorable, especially for smooth non-convex objects, when there exist inter-reflections on the NLOS object. This makes separating three-bounce light path from higher-bounce light paths harder. However, in this case, for Lambertian reflectance, higher order bounce light paths attenuate faster and it is easier to identify three-bounce light paths. Specular BRDF is also not suitable for non-smooth object since some lighting/sensing pairs cannot receive any photon.

A subset of available measurements. The ToF of first-returning photons is a subset of the measurements encoded in the full 5D light transient. If the 5D light transient is available, all the measurements should be used to provide additional robustness especially when imaging complex scenes.

4.8 Conclusion

Non-line-of-sight imaging uses properties measured in the field of view to infer the properties of the invisible parts of the scene. This is done by considering the attributes of three-bounce light paths. The time-of-flight corresponds to the path length constraints and the radiance corresponds to the reflectance information of the NLOS scene. The main difficulty for estimating the shape of non-line-of-sight imaging is that different parts of the object contribute to the measurements. This means that to estimate the shape, we not only need to deal with the entanglement of geometry and reflectance, but also need to unmix the contribution from different parts of the object.

By considering how light propagates the scene, we find the first-returning photons to carry the geometric information of the hidden scene. Our finding is in terms of three observations that includes what kind of scene cannot possibly generate the observed first-returning photons, which leads to the space carving algorithm for NLOS imaging. Another important key is first-returning photon follows Fermat’s principle. By assuming local planarity, we derive an algorithm to localize NLOS scene points in 3D and estimate their surface normals. The proposed method proposed is computationally lightweight, hence is suitable for initializing and accelerating more complex methods that utilize the full light transient.

Contributions In particular, we make several contributions:

- *Acquisition of first-returning photons.* We provide theoretical and experimental justifications regarding the ability to measure first-returning photons using a SPAD. We highlight how sensing first-returning photons helps simplify acquisition, for instance, shortening the integration time needed.
- *Derivation of geometric properties of first-returning photons.* We establish several geometric properties of the light path traced by first-returning photons. One of the most important properties is that the ToF of the first-returning photon is a function of the geometry of the NLOS object. This means that a shape estimation framework based on first-returning photon is invariant to the NLOS object’s reflectance.
- *Design of algorithms for NLOS shape recovery using first-returning photons.* We derive constraints induced by the ToF of the first-returning photon and subsequently provide three algorithms to recover the hidden object:
 1. *Space carving.* The ToF of the first-returning photon implies that the closest NLOS points are a specific distance away from the LOS scene points. Based on this observation, we derive a space carving algorithm that restricts the spatial extent of the NLOS object.
 2. *Shape and normal from first-returning photons using local planarity constraints.* We derive constraints that relate the 3D positions of the NLOS scene points to the ToF of first-returning photons. This enables a simple algorithm for estimating the 3D locations of the NLOS scene points. Once the scene point is localized in space, we show that the surface normal can also be recovered provided the NLOS scene is locally smooth. This derivation relies on Fermat’s principle of shortest paths.
 3. *Shape and normal from first-returning photons using a source-separation technique.* We show that the ToF of first-returning photons from a planar object is a quadratic function of

space. We use this observation to cluster first-returning photons from the same (locally) planar object and to robustly recover its plane parameters.

- *NLOS object tracking using first-returning photon.* Since first-returning photons are easy to acquire, we can track a moving NLOS object at a high frame rate by measuring a small number of first-returning photons.
- *Creation of an NLOS dataset.* We create an NLOS imaging dataset collected with a SPAD sensor along with the scene descriptions. This new dataset includes extensive real experiments with ground truth information for a comparison of different algorithms.

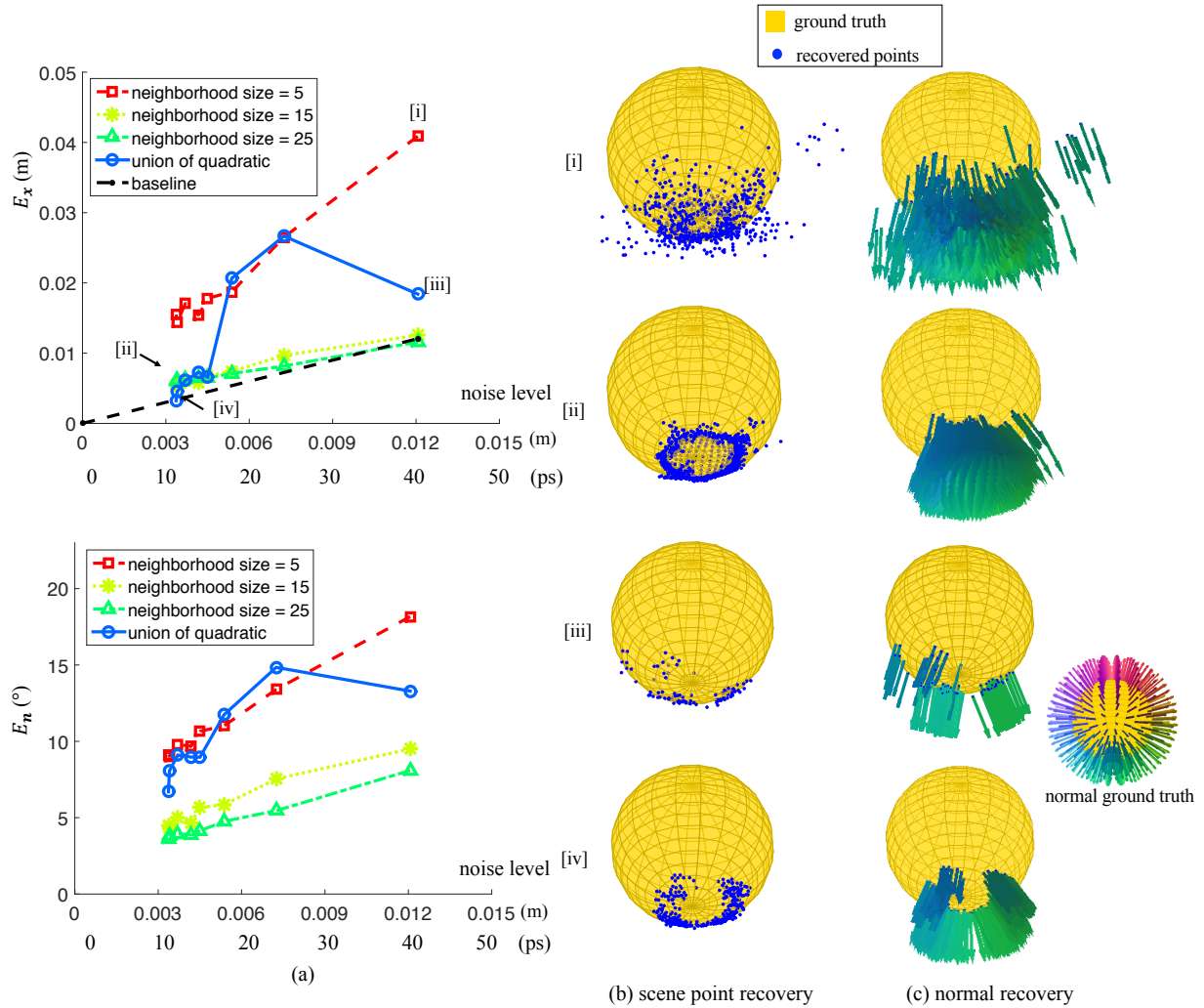


Figure 4.22: **Shape recovery with different noise levels.** We are able to vary the noise level by varying the number of photons used in rendering. (a) The shape recovery error with respect to different noise levels. [i] Recovery result of noise level at 0.0123 m with neighborhood size = 5 with Alg 4. [ii] Recovery result of noise level at 0.0034 m with neighborhood size = 15 with Alg 4. [iii] Recovery result of noise level at 0.0123 m with Alg 5. [iv] Recovery result of noise level at 0.0034 m with Alg 5.

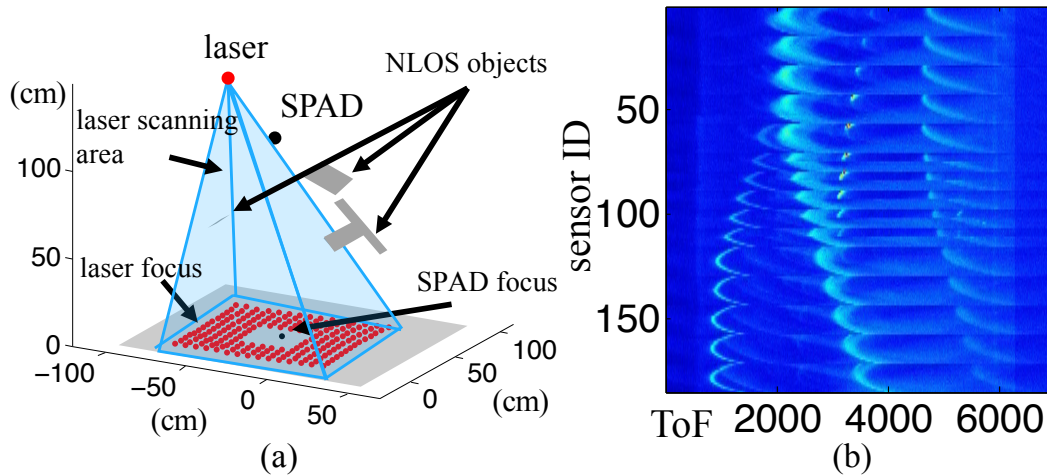


Figure 4.23: **NLOS imaging setup in [15]**. (a) The laser scans 185 locations on the wall while SPAD focuses on one spot on the wall. (b) The 5D light transient collected by [15]. We extract first-returning photons from (b) to demonstrate the effectiveness of our proposed algorithms.

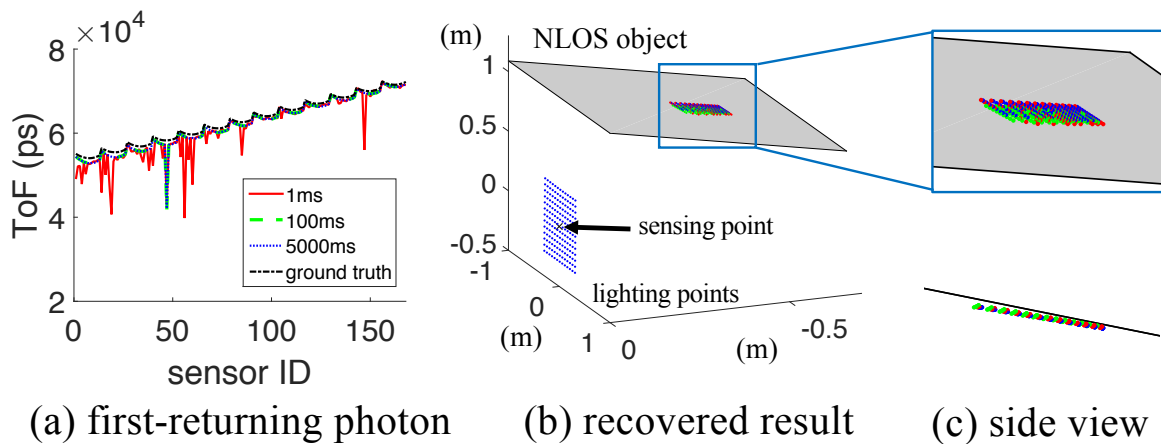


Figure 4.24: **NLOS shape recovery with different integration time**. We collect light transient data of a plane using different integration time. (a) We can see that the location of the first-returning photon is roughly invariant with respect to the integration time. Since we also measured the ground truth location of the planar NLOS scene, we can compute the theoretical ground truth position of the first returning photon. (b,c) We show the recovered results in two different angles.

#	Alg 4		Alg 5	
	E_x (m)	E_n ($^\circ$)	E_x (m)	E_n ($^\circ$)
1	0.120	71.619	0.148	73.165
2	0.024	15.222	0.009	19.496
3	0.019	13.919	0.002	8.966
4	0.016	2.544	0.021	0.507
5	0.009	2.147	0.014	0.652
6	0.012	3.594	0.006	14.192
7	0.087	40.291	0.013	19.288
8*	0.034	15.406	0.014	1.041
9*	0.030	6.279	0.027	3.122

Table 4.2: NLOS shape recovery results on dataset in [102].

#	intercept (m)	angle ($^\circ$)	Alg 4		Alg 5	
			E_x (m)	E_n ($^\circ$)	E_x (m)	E_n ($^\circ$)
1	1.09	15	0.047	34.887	0.020	31.992
2	1.09	30	0.016	16.973	0.016	22.153
3*	1.09	45	0.028	4.757	0.027	4.775
4*	1.09	60	0.043	6.921	0.040	5.007
5*	1.09	75	0.040	6.211	0.034	5.481
6	1.09	90	0.110	26.450	0.019	12.810
7	1.09	105	0.448	62.930	0.2608	48.643
8	0.30	60	0.038	3.166	0.039	3.213
9	0.46	60	0.090	7.072	0.088	6.558
10	0.66	60	0.049	3.514	0.051	3.102

Table 4.3: Recovery results with planes at different positions.

#	Alg 4		Alg 5	
	E_x (m)	E_n ($^\circ$)	E_x (m)	E_n ($^\circ$)
1	0.037	11.409	0.041	14.530
2	0.024	14.253	0.028	16.809
3	0.047	13.913	0.051	15.997
4	0.067	43.834	0.047	37.930

Table 4.4: Recovery results on curved objects.

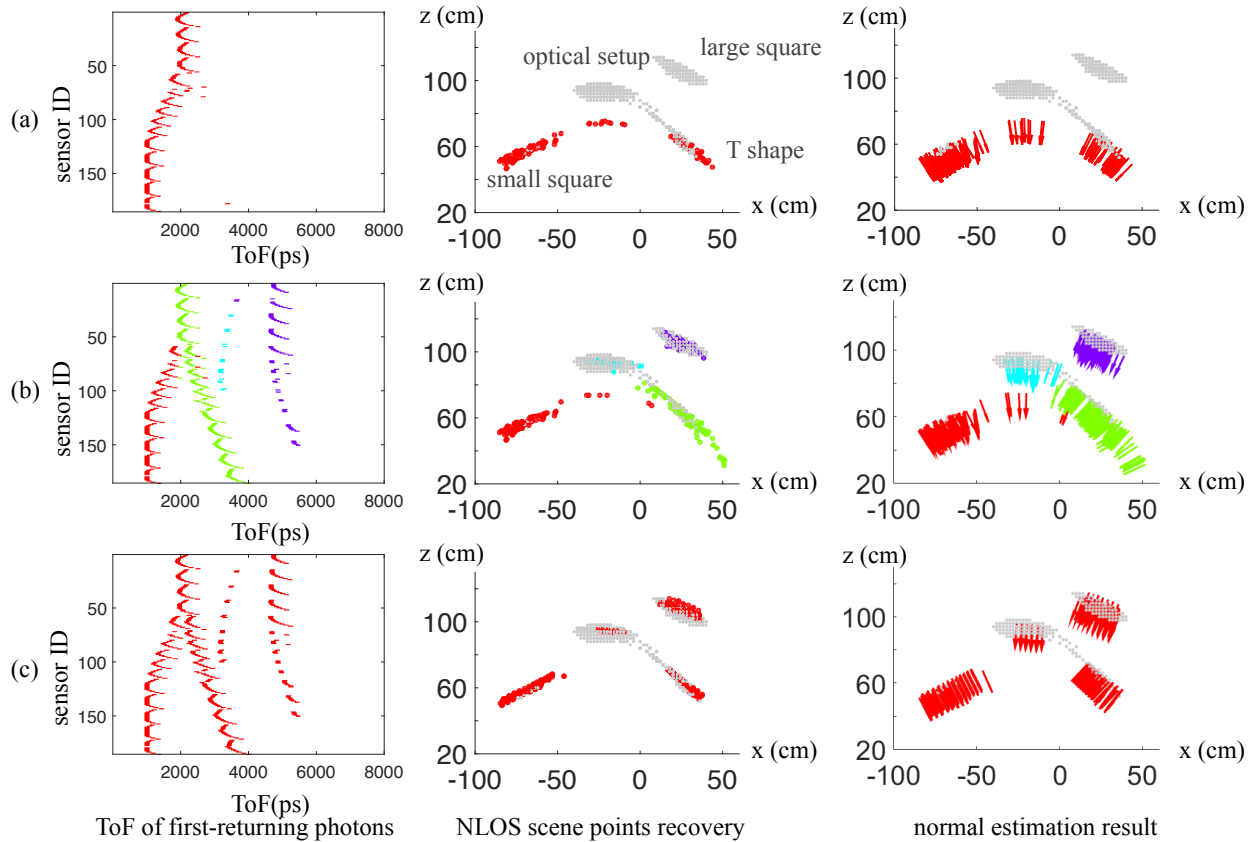


Figure 4.25: **NLOS shape estimation from first-returning photons collected by SPAD [15]**. We compare our proposed method using solely first-returning photons (red and other color dots) with a elliptical tomography method using the full 5D light transient (grey dots). (a) First-returning photons by finding where the photon count increases sharply in the light transient. We show that we can recover points closer to the wall and the recovered points align with the result using a elliptical tomography method. (b) We use signal processing to extract the first-returning photon from different NLOS surfaces. For each surface, we repeat the process in (a). We can see that by finding first-returning photons from different surfaces in the 5D light transient, we can recover most of the NLOS scene. (c) We collect the first-returning photon by finding the peak in the light transient and directly apply the union of quadratic models to automatically cluster measurements that belong to the same surface. Note that this can help eliminating the artifact caused by clustering wrong subsets of measurements.

#	scene	Alg 4		Alg 5	
		E_x (m)	E_n ($^\circ$)	E_x (m)	E_n ($^\circ$)
1	2 planes	0.026	10.813	0.036	7.592
2	2 planes	0.014	10.900	0.011	7.115
3	3 planes	0.029	12.307	0.008	5.629
4	3 planes	0.029	12.676	0.017	10.649
5	1 plane, 1 ball	0.058	25.967	0.014	4.234
6	2 balls	0.104	52.065	0.060	41.343

Table 4.5: Recovery results on multi-object scene

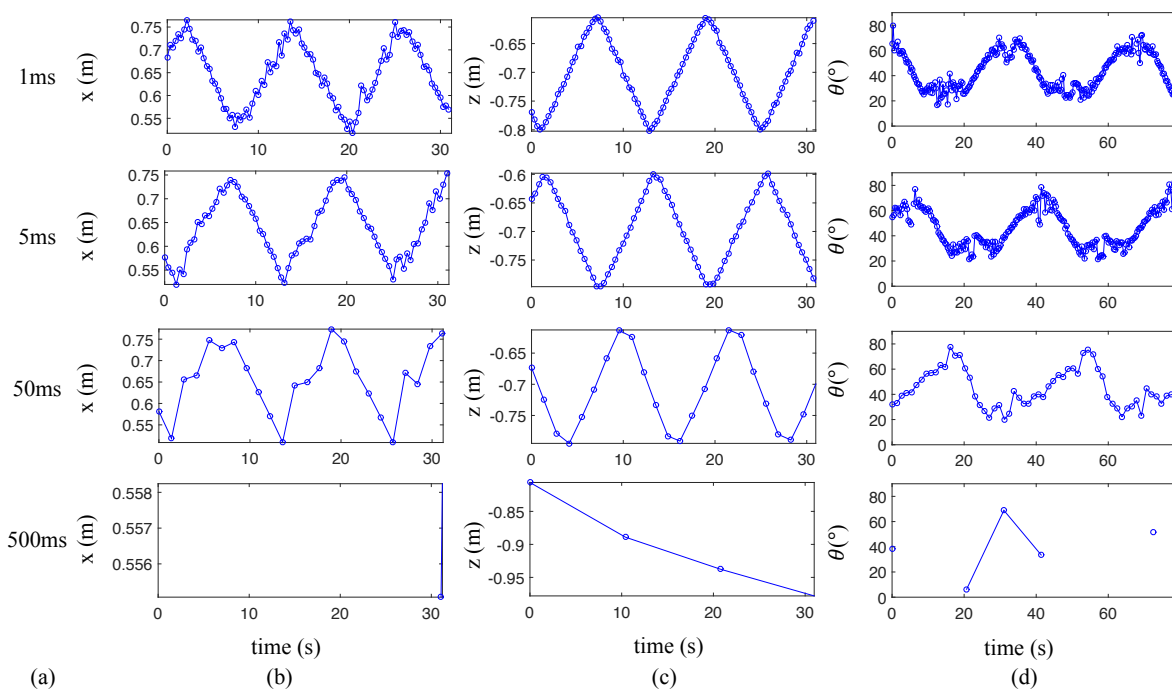


Figure 4.26: **NLOS object tracking using a SPAD.** Our proposed method of using first-returning photon along with Alg. 5 that utilizes quadratic modeling of planes enable us to use few pairs of illumination and sensing points and short integration time to find parameters of a plane. (a) We show the result for tracking of 4 different integration time. We measure 20 pairs of illumination and sensing points to estimate one plane. (b) The NLOS object moves back and forth along x direction, which is parallel to the LOS wall. (c) The NLOS object moves back and forth along z direction, which is perpendicular to the wall. (d) The NLOS object rotates from parallel to the wall to perpendicular to the wall.

Chapter 5

Non-line-of-sight imaging — Shape from radiance of three-bounce light paths

We showed in Chapter 4 that by identifying the physical laws governing subset of the measurements of the 5D light transient, there exist some simple yet powerful tools to deal with the hard shape estimation problem that involves three-bounce light paths. The three algorithms are computationally lightweight and easy to implement. However, as mentioned in Section 4.7, since the scene points that generate first-returning photons are completely determined by the shape of the object and the imaging setup, we can only reconstruct partial information regarding the hidden scene. Still, for algorithms mentioned in Chapter 4, we are only using the ToFs of the first-returning photons. There is a lot more information since the imaging system typically collects the full 5D light transient information. Therefore, in this chapter, we detail another technique that make use of the whole available information and in conjunction with results from Chapter 4 creates a shape estimation quality that is orders of magnitude better than the state-of-the-art in the NLOS imaging field.

We revisit the image formation model of time-resolved measurements used by most existing techniques in NLOS imaging field which is introduced by Velten *et al.* [135]. This model represents the NLOS scene as a so-called albedo volume, where each voxel acts as an isotropic reflector with an associated albedo value. The volume representation allows approximately formulating transient light transport in the NLOS scene with only linear algebraic operations. In turn, the unknown albedo volume can be recovered from the transient measurements by solving a, potentially regularized, linear least-squares system [5, 45, 49, 50, 96].

This mathematical tractability comes at the cost of physical accuracy. In NLOS scenes consisting of opaque objects, light transport is the result of discrete light-surface interactions at the object interfaces, rather than continuous light-volume interactions. Additionally, these interactions include effects such as normal-dependent shading and non-Lambertian reflectance, which are ignored by the albedo volume. On the other hand, instead of volumetric albedo, given a representation of the NLOS objects' surface and reflectance, light transport can be modeled exactly using the rendering equation [62]. However, unlike the albedo volume model, evaluating this equation is

only possible through computationally expensive Monte Carlo rendering operations [27, 106, 131]. This increased computational complexity has so far hindered the adoption of the rendering equation in NLOS reconstruction techniques.

We first show expressions for the radiometric measurements captured as a function of properties of the NLOS scene. To help navigate this section, we use the notation shown in Figure 5.1(a) to write down expressions for the radiometric measurements captured under this setting as a function of the NLOS scene properties. The image formation model expressions are common-place in the physics-based rendering literature (see, for instance, [27, 106, 131]), but we describe them in detail as the necessary background for deriving the algorithm for shape recovery.

Image formation model. We restrict our attention to light effects from three-bounce paths of the form $l_0 \rightarrow l \rightarrow x \rightarrow s \rightarrow s_0$ where $x \in \mathcal{S}_{\text{NLOS}}$; that is, paths that, between the virtual source l and virtual detector s , have a single interaction with the NLOS surface at a point $x \in \mathcal{S}_{\text{NLOS}}$. We make this simplification motivated from previous observations that photons following higher-order paths are difficult to detect with existing sensors [102]. We additionally ignore light following direct paths without interacting with the NLOS surface $\mathcal{S}_{\text{NLOS}}$, as this light component is typically removed using time-gating mechanisms [15]. Additionally, for each pair of virtual points l and s , we assume that we have calibrated our measurements so that we can neglect the radiometric and pathlength terms for the connections $l_0 \rightarrow l$ and $s \rightarrow s_0$.

Under these assumptions, we can use the path integral formulation of light transport [131] to write the intensity measured by the sensor s_0 at time t as

$$I(t; l, s) = \int_{\mathcal{S}_{\text{NLOS}}} \overbrace{W(x; t) f(x, \hat{n}(x))}^{g(x, \hat{n}(x))} \cdot V(x, l) V(x, s) dA(x), \quad (5.1)$$

where $A(x)$ is the area measure on $\mathcal{S}_{\text{NLOS}}$, \hat{n} is the normal of a surface at a specific point, and $W(x; t)$, $f(x, \hat{n}(x))$, $V(x, s)$ will be discussed below. We note that, because of the three-bounce assumption, the usual path integral reduces to a single surface integral over the NLOS surface $\mathcal{S}_{\text{NLOS}}$.

The *radiometric throughput* f in Equation (5.1) is the radiance that flows through the path $l \rightarrow x \rightarrow s$,

$$f(x, \hat{n}(x)) = f_s(\hat{n}(x), \hat{\omega}_l(x), \hat{\omega}_s(x)) \cdot \frac{\langle -\hat{\omega}_l(x), \hat{n}(l) \rangle \langle \hat{\omega}_l(x), \hat{n}(x) \rangle}{\|x - l\|^2} \cdot \frac{\langle -\hat{\omega}_s(x), \hat{n}(s) \rangle \langle \hat{\omega}_s(x), \hat{n}(x) \rangle}{\|x - s\|^2}, \quad (5.2)$$

where f_s is the BRDF of $\mathcal{S}_{\text{NLOS}}$ at point x , $\hat{\omega}_l(x)$ is the normalized vector parallel to $l - x$, and likewise for $\hat{\omega}_s(x)$.

The *temporal importance* W models the mechanism by which the sensor selects paths of length within some specific range for each measurement $I(t; \mathbf{l}, \mathbf{s})$ ¹,

$$W(\mathbf{x}, t) = \text{rect} \left(\frac{\tau(\mathbf{x}) - t}{T} \right),$$

where rect is the unit rectangular function, T is the sensor’s temporal resolution, and τ is the length of path $\mathbf{l} \rightarrow \mathbf{x} \rightarrow \mathbf{s}$,

$$\tau(\mathbf{x}) = \|\mathbf{x} - \mathbf{l}\| + \|\mathbf{x} - \mathbf{s}\|.$$

Finally, the *visibility function* V is a binary indicator of occlusion between two points,

$$V(\mathbf{x}, \mathbf{y}) = \begin{cases} 1, & \mathbf{x}, \mathbf{y} \text{ are visible to each other,} \\ 0, & \text{otherwise.} \end{cases}$$

Comparison to the albedo volume model. It is instructive to compare the surface integral formulation of Equation (5.1) with the albedo volume model of Velten et al. [135]. This model represents the NLOS scene as an albedo function $\rho(\mathbf{x})$, defined on all points of a continuous three-dimensional volume, $\mathbf{x} \in \mathcal{V}_{\text{NLOS}}$. Then, transient measurements are expressed as a volume integral,

$$I(t; \mathbf{l}, \mathbf{s}) = \int_{\mathcal{V}_{\text{NLOS}}} \frac{W(\mathbf{x}; t) \rho(\mathbf{x})}{\|\mathbf{x} - \mathbf{l}\|^2 \|\mathbf{x} - \mathbf{s}\|^2} dV(\mathbf{x}), \quad (5.3)$$

where $V(\mathbf{x})$ is the standard measure on $\mathcal{V}_{\text{NLOS}}$. Compared to Equation (5.1), we note that the integrand of Equation (5.3) constrains the reflectance function f_s to be Lambertian and omits the normal-related shading terms and the visibility terms V . Recent extensions incorporate normal and visibility effects through additional volumetric functions defined everywhere on $\mathcal{V}_{\text{NLOS}}$ [49]. However, the albedo, normal, and visibility volumetric functions are treated as independent of each other, even though they are in fact intertwined as functions of the underlying NLOS surface.

Despite the lack of physical accuracy, the albedo volume model is attractive because of its mathematical convenience: Through a straightforward discretization of the volume integral of Equation (5.3), forward evaluations of the model become simple matrix-vector multiplication operations. Consequently, inverting the model to reconstruct the NLOS scene can be posed as a linear least-squares problem. By contrast, forward evaluations of the surface integral of Equation (5.1) rely on involved surface quadrature methods, or Monte Carlo rendering. This makes inverting the model for NLOS reconstruction non-trivial.

In this chapter, we show a NLOS imaging method that directly deals with the physically accurate image formation model [128]. We overcome the computational complexity and introduce a pipeline that reconstructs NLOS object shape, in the form of a triangular mesh, and complex reflectance, in the form of a microfacet BRDF, while accurately taking into account the underlying light transport physics. The overview of the surface optimization pipeline is shown in Figure 5.1.

¹We treat geometric pathlength and time of flight as equivalent, with the understanding that they relate to each other through the speed of light.

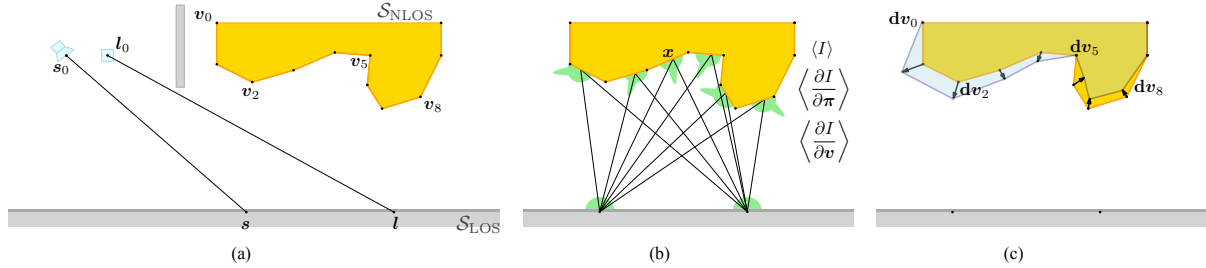


Figure 5.1: **NLOS surface optimization pipeline.** (a) 2D visualization and notation. (b) We sample points x on the surface to compute the estimated transient and gradient with respect to reflectance π and vertices v of the triangular mesh. (c) From the computed gradients, we update the surface.

At the core of our pipeline is a differentiable formulation of the rendering equation in the NLOS setting. This formulation enables the use of Monte Carlo rendering to efficiently estimate derivatives of radiometric measurements with respect to shape and reflectance parameters. We combine an optimized differentiable rendering implementation with stochastic optimization in an inverse rendering framework [85], where we iteratively deform an NLOS surface so as to minimize the difference between measured and rendered light transients. We augment this surface optimization pipeline with geometric processing tools that help improve the quality of the resulting triangular mesh. Through experiments on synthetic and measured data, we show that this pipeline can produce NLOS surface reconstructions at a level of detail comparable to what is achieved by albedo-volume methods using two orders of magnitude more measurements, while additionally recovering non-Lambertian reflectance.

5.1 Related work

In this chapter, we deal with the NLOS imaging problem with a new perspective. We discussed prior works in NLOS imaging in Chapter 4. Here we discuss two additional fields that we rely on for our surface optimization pipeline based on inverse rendering.

Surface optimization. Surface optimization is a classical approach for 3D reconstruction in computer vision, where it is commonly applied for image-based reconstruction [4, 22, 23, 32, 73, 109, 114, 119, 146, 147]. In a related context, surface optimization techniques are used in mesh editing applications for computer graphics [24, 29, 79, 87, 120]. At a high-level, both types of applications operate by first defining an objective function (or energy) as an integral on a surface. Then, they derive expressions for the derivatives of this surface integral with respect to some surface representations. Finally, these derivatives are used to create a flow process that progressively deforms some initial surface, until the objective function is minimized. The derivative expressions typically relies on tools from differential geometry, and has been demonstrated for both implicit (e.g., level sets [92]) and parametric (e.g., triangular meshes [23]) surface representations. Similar surface integrals arise in the context of NLOS imaging through the rendering equation. Therefore, we take

advantage of this mathematical machinery to perform surface optimization for NLOS reconstruction.

Differentiable rendering. Differentiable rendering has been introduced as a methodology for recovering physical unknowns from image measurements, which can include direct-only [83] and global illumination effects (e.g., scattering [36, 38, 39, 67, 76, 149], or interreflections [82]). Typically, differentiable rendering is used to perform analysis-by-synthesis, also known as inverse rendering [85, 100, 101], which refers to searching for values of physical parameters that, when used to synthesize images, can reproduce input images accurately. To efficiently perform this search through gradient-descent optimization, differentiable rendering is used to estimate derivatives of images with respect to the unknown parameters. In our setting, we devise new differentiable rendering algorithms that are compatible with the surface optimization problem, and that are tailored to the NLOS image formation model for increased estimation efficiency.

5.2 Problem statement

We are given a set of calibrated transient measurements $\{\tilde{I}_m(t), m = 1, \dots, M\}$, corresponding to pairs of virtual points $\{(\mathbf{l}_m, \mathbf{s}_m), m = 1, \dots, M\}$. We additionally adopt parametric forms $\mathcal{S}_{\text{NLOS}}[\mathbf{v}]$ and $f_s[\boldsymbol{\pi}]$ for the NLOS surface and reflectance, respectively.

The goal of this chapter is to search for the parameter values, \mathbf{v} and $\boldsymbol{\pi}$, that can be used to simulate transients that best match our measurements. Formally, we minimize the following energy function,

$$E(\mathbf{v}, \boldsymbol{\pi}) = \frac{1}{2} \sum_{m,t} \|\tilde{I}_m(t) - I[\mathbf{v}, \boldsymbol{\pi}](t; \mathbf{l}_m, \mathbf{s}_m)\|^2. \quad (5.4)$$

We use the notation $I[\mathbf{v}, \boldsymbol{\pi}](t; \mathbf{l}, \mathbf{s})$ to indicate that a rendered transient is a function, through Equation (5.1), of the surface and reflectance parameters \mathbf{v} and $\boldsymbol{\pi}$.

5.3 Analysis by synthesis optimization

We aim to use gradient-descent optimization, in order to efficiently minimize the analysis-by-synthesis objective and recover the NLOS surface and reflectance parameters. Differentiating the loss function $E(\mathbf{v}, \boldsymbol{\pi})$ of Equation (5.4) with respect to surface and reflectance parameters, we obtain

$$\frac{\partial E}{\partial \mathbf{y}} = - \sum_{m,t} (\tilde{I}_m(t) - I(t; \mathbf{l}_m, \mathbf{s}_m)) \frac{\partial I(t; \mathbf{l}_m, \mathbf{s}_m)}{\partial \mathbf{y}}, \quad (5.5)$$

where \mathbf{y} can be either \mathbf{v} or $\boldsymbol{\pi}$. Evaluating the derivatives requires computing not only the transients I , but also their derivatives $\partial I / \partial \boldsymbol{\pi}$ and $\partial I / \partial \mathbf{v}$ with respect to reflectance and surface parameters. This is challenging because I is not an analytical function of these parameters, but is only related to them through the surface integral of Equation (5.1).

We overcome this obstacle using an approach based on differentiable rendering. We prove that the derivatives $\partial I/\partial \pi$ and $\partial I/\partial v$ can be expressed as surface integrals analogous to that of Equation (5.1). This allows us to derive efficient Monte Carlo rendering algorithms for stochastically approximating the reflectance and surface derivatives. We can, then, combine these stochastic estimates with stochastic gradient descent optimization [68] to minimize Equation (5.4). In the rest of this section, we first describe our choices for NLOS surface and reflectance parameterization, then provide an overview of our approach differentiable rendering approach.

5.3.1 Differentiating transients

Surface parameterization. We represent the NLOS surface $\mathcal{S}_{\text{NLOS}}$ as a triangular mesh with boundary, which we represent using two matrices: First, a $3 \times V$ geometry matrix \mathbf{V} providing the three-dimensional coordinates of its V vertices. Second, a $3 \times T$ topology matrix \mathbf{T} providing the integer vertex indices of its T triangles. We do not assign any normal or texture parameters to the vertices, and at every point on the mesh, we assume that the surface normal is equal to the corresponding triangle’s face normal. We use meshes instead of an implicit representation (e.g., level sets [92] or signed distance functions [21] to facilitate efficient Monte Carlo rendering (see Section 5.3.2). On the other hand, this complicates optimization due to the need to handle the discrete topology matrix \mathbf{T} . As is common in mesh optimization, we use differentiable rendering to minimize Equation (5.4) only with respect to mesh vertices. During this iterative minimization, we use standard geometry processing tools to improve the mesh topology (Section 5.4).

Reflectance parameterization. We assume that the NLOS surface has a spatially-uniform BRDF, which we represent using the widely-adopted GGX microfacet BRDF, as described by Walter et al. [139]. We first define the surface normal $\hat{\mathbf{n}}$, incoming and outgoing directions ω_i and ω_o , respectively, and the half-vector $\mathbf{h} = (\omega_i + \omega_o)/\|\omega_i + \omega_o\|$. Then, the BRDF model can be written as

$$\begin{aligned} f_s[\pi](\hat{\mathbf{n}}, \omega_i, \omega_o) &= f_{\text{GGX}}[\alpha](\hat{\mathbf{n}}, \omega_i, \omega_o) \\ &= \frac{F_{\text{Cook-Torrance}}(\omega_o, \mathbf{h}) \cdot D_{\text{GGX}}(\hat{\mathbf{n}}, \mathbf{h}) \cdot G_{\text{Smith}}(\hat{\mathbf{n}}, \omega_i, \omega_o)}{4}, \end{aligned} \quad (5.6)$$

where

$$D_{\text{GGX}}(\hat{\mathbf{n}}, \mathbf{h}) = \frac{\alpha^2}{\pi \left[\langle \hat{\mathbf{n}}, \mathbf{h} \rangle^2 (\alpha^2 - 1) + 1 \right]^2}, \quad (5.7)$$

$$G_{\text{Smith}}(\hat{\mathbf{n}}, \omega_i, \omega_o) = G_1(\hat{\mathbf{n}}, \omega_i) \cdot G_1(\hat{\mathbf{n}}, \omega_o), \quad (5.8)$$

$$G_1(\hat{\mathbf{n}}, \omega) = \frac{2}{\langle \hat{\mathbf{n}}, \omega \rangle + \sqrt{\alpha^2 + (1 - \alpha^2) \langle \hat{\mathbf{n}}, \omega \rangle^2}}, \quad (5.9)$$

and $F_{\text{Cook-Torrance}}(\omega_o, \mathbf{h})$ is the Fresnel reflection function for an ideal reflector, independent of α . This results in a one-dimensional BRDF parameterization $\pi = [\alpha]$, where the parameter $\alpha \in [0, 1]$ controls the roughness of the specular lobe.

Derivatives as surface integrals. We now state the main technical result of this chapter, which allows us to derive expressions for the derivatives of the image formation model (5.1) with respect to surface geometry and reflectance.

Proposition 3. *The derivatives of a transient $I(t; \mathbf{l}, \mathbf{s})$ with respect to reflectance and mesh vertices can be written as:*

$$\frac{\partial I}{\partial \mathbf{v}} = \int_{S_{\text{NLOS}}} g_s(\mathbf{x}, \hat{\mathbf{n}}(\mathbf{x})) V(\mathbf{x}, \mathbf{l}) V(\mathbf{x}, \mathbf{s}) dA(\mathbf{x}), \quad (5.10)$$

$$\frac{\partial I}{\partial \boldsymbol{\pi}} = \int_{S_{\text{NLOS}}} g_r(\mathbf{x}, \hat{\mathbf{n}}(\mathbf{x})) V(\mathbf{x}, \mathbf{l}) V(\mathbf{x}, \mathbf{s}) dA(\mathbf{x}), \quad (5.11)$$

for appropriate functions g_s and g_r .

In the following, we provide detailed expressions for g_s and g_r .

Reflectance derivatives. Differentiating the surface integral of Equation (5.1) with respect to reflectance is straightforward. By observing that the integration measure is independent of the reflectance parameters $\boldsymbol{\pi}$, and under weak continuity conditions that are known to be satisfied for radiometric integrals [6], we can simply exchange the order of differentiation and integration to obtain

$$\left. \frac{\partial I(t; \mathbf{l}, \mathbf{s})}{\partial \boldsymbol{\pi}} \right|_{\mathbf{v}, \boldsymbol{\pi}} = \int_{S_{\text{NLOS}}} \overbrace{W(\mathbf{x}; t) \frac{\partial f(\mathbf{x}, \hat{\mathbf{n}}(\mathbf{x}))}{\partial \boldsymbol{\pi}} \Big|_{\mathbf{v}, \boldsymbol{\pi}}}^{g_r(\mathbf{x}, \hat{\mathbf{n}}(\mathbf{x}))} \cdot V(\mathbf{x}, \mathbf{l}) V(\mathbf{x}, \mathbf{s}) dA(\mathbf{x}),$$

where

$$\begin{aligned} \left. \frac{\partial f(\mathbf{x}, \hat{\mathbf{n}}(\mathbf{x}))}{\partial \boldsymbol{\pi}} \right|_{\mathbf{v}, \boldsymbol{\pi}} &= \left. \frac{\partial f_s[\boldsymbol{\pi}](\hat{\mathbf{n}}(\mathbf{x}), \hat{\boldsymbol{\omega}}_l(\mathbf{x}), \hat{\boldsymbol{\omega}}_s(\mathbf{x}))}{\partial \boldsymbol{\pi}} \right|_{\mathbf{v}, \boldsymbol{\pi}} \\ &\cdot \frac{\langle -\hat{\boldsymbol{\omega}}_l(\mathbf{x}), \hat{\mathbf{n}}(\mathbf{l}) \rangle \langle \hat{\boldsymbol{\omega}}_l(\mathbf{x}), \hat{\mathbf{n}}(\mathbf{x}) \rangle}{\|\mathbf{x} - \mathbf{l}\|^2} \\ &\cdot \frac{\langle -\hat{\boldsymbol{\omega}}_s(\mathbf{x}), \hat{\mathbf{n}}(\mathbf{s}) \rangle \langle \hat{\boldsymbol{\omega}}_s(\mathbf{x}), \hat{\mathbf{n}}(\mathbf{x}) \rangle}{\|\mathbf{x} - \mathbf{s}\|^2}, \end{aligned}$$

and the derivatives $\partial f_s / \partial \boldsymbol{\pi}$ can be computed analytically from Equations (5.6)-(5.9),

$$\frac{\partial f_s}{\partial \alpha} = \frac{\partial f_{\text{GGX}}[\alpha](\hat{\mathbf{n}}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o)}{\partial \alpha}.$$

We used symbolic differentiation to compute this derivative, and then implemented the resulting expression in our rendering code.

Surface derivatives. In contrast to the reflectance case, differentiating the surface integral of Equation (5.1) with respect to mesh vertices is complicated by the fact that the integration measure $A(\mathbf{x})$ is now a function of these same mesh vertices. We tackle this by using recent results on

analytically expressing gradient flows from mesh functionals as surface integrals [28, 29]. These results have also been used by Delaunoy and Prados [23] for surface optimization in line-of-sight reconstruction problems (e.g., shape from shading).

An additional complicating factor is the presence of the binary and discontinuous visibility terms V in the integrand. We can ameliorate the second problem by making the approximation that the visibility terms are independent of the mesh geometry. This approximation is justified by the fact that the visibility terms have non-zero derivatives only on a zero-measure part of the surface (that is, on occluding contours [72]), and is common in computer vision and graphics problems [4, 79, 114]. Delaunoy and Prados [23] showed that differentiation is possible even without approximating the visibility terms V as constant. We can similarly extend Proposition 3 to account for visibility. However, in practice we found that this complicates Monte Carlo rendering without significantly improving the optimization results.

We derive Equation (5.10) with the approximation that the visibility terms are independent of the mesh geometry.

$$\begin{aligned} & \left. \frac{\partial I(t; \mathbf{l}, \mathbf{s})}{\partial \mathbf{v}_i} \right|_{\mathbf{v}, \pi} \\ &= \sum_{k \in J_i} \int_{\mathcal{T}_k} \overbrace{[g_{s1}(\mathbf{x}, \hat{\mathbf{n}}(\mathbf{x}); i, k) + g_{s2}(\mathbf{x}, \hat{\mathbf{n}}(\mathbf{x}); i, k)]}^{g_s(\mathbf{x}, \hat{\mathbf{n}}(\mathbf{x}); i, k)} \\ & \quad \cdot V(\mathbf{x}, \mathbf{l}) V(\mathbf{x}, \mathbf{s}) \, dA(\mathbf{x}), \end{aligned}$$

where

$$\begin{aligned} g_{s1}(\mathbf{x}, \hat{\mathbf{n}}(\mathbf{x}); i, k) &= \nabla_{\mathbf{x}} g(\mathbf{x}, \hat{\mathbf{n}}_k) \phi_i(\mathbf{x}), \\ g_{s2}(\mathbf{x}, \hat{\mathbf{n}}(\mathbf{x}); i, k) &= -\frac{\mathbf{e}_{k,i}}{2A_k} \wedge [g(\mathbf{x}, \hat{\mathbf{n}}_k) \hat{\mathbf{n}}_k + g_{\hat{\mathbf{n}}}(\mathbf{x}, \hat{\mathbf{n}}_k)]. \end{aligned}$$

J_k corresponds to the set of faces that contain vertex i . $\nabla_{\mathbf{x}}$ is the gradient of g with respect to \mathbf{x} , $\phi_i(\mathbf{x})$ is the linear interpolating basis function. $\mathbf{e}_{k,i}$ is the opposite edge of vertex i in the face k pointing counterclockwise. Operator \wedge denotes the cross product. Finally,

$$g_{\hat{\mathbf{n}}} = \nabla_{\hat{\mathbf{n}}} g(\mathbf{x}, \hat{\mathbf{n}}_k) - \langle \nabla_{\hat{\mathbf{n}}} g(\mathbf{x}, \hat{\mathbf{n}}_k), \hat{\mathbf{n}}_k \rangle \hat{\mathbf{n}}_k.$$

In the following, we detail some quantities that are used in the above expressions. For simplicity, we show the expression for the case of confocal imaging ($\mathbf{l} = \mathbf{s}$). The derivation for the non-confocal case is straightforward, only involves longer expressions. As defined in Equations

(5.1) and (5.2),

$$\begin{aligned}
g(\mathbf{x}, \hat{\mathbf{n}}(\mathbf{x})) &= W(\mathbf{x}; t) f(\mathbf{x}, \hat{\mathbf{n}}(\mathbf{x})) \\
&= W(\mathbf{x}; t) f_s(\hat{\mathbf{n}}(\mathbf{x}), \hat{\omega}_l(\mathbf{x}), \hat{\omega}_s(\mathbf{x})) \\
&\quad \cdot \frac{\langle -\hat{\omega}_l(\mathbf{x}), \hat{\mathbf{n}}(\mathbf{l}) \rangle \langle \hat{\omega}_l(\mathbf{x}), \hat{\mathbf{n}}(\mathbf{x}) \rangle}{\|\mathbf{x} - \mathbf{l}\|^2} \\
&\quad \cdot \frac{\langle -\hat{\omega}_s(\mathbf{x}), \hat{\mathbf{n}}(\mathbf{s}) \rangle \langle \hat{\omega}_s(\mathbf{x}), \hat{\mathbf{n}}(\mathbf{x}) \rangle}{\|\mathbf{x} - \mathbf{s}\|^2}, \\
&= W(\mathbf{x}; t) f_s(\hat{\mathbf{n}}(\mathbf{x}), \hat{\omega}_l(\mathbf{x}), \hat{\omega}_s(\mathbf{x})) g_0(\mathbf{x}, \hat{\mathbf{n}}(\mathbf{x}))
\end{aligned}$$

The temporal importance W function is a unit rectangular function. This means that the derivatation will only be non-zero on the boundary. To make g defferentiable, we substitute W with a Gaussian that has full width at half max equal to the temporal resolution of the sensor. Then,

$$\begin{aligned}
W(\mathbf{x}; t) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\tau(\mathbf{x}) - t)^2}{2\sigma^2}\right) \\
\frac{\partial W(\mathbf{x}; t)}{\partial \mathbf{x}} &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\tau(\mathbf{x}) - t)^2}{2\sigma^2}\right) \frac{2[\tau(\mathbf{x}) - t]}{\sigma^2} \hat{\omega}_l \\
\frac{\partial W(\mathbf{x}; t)}{\partial \hat{\mathbf{n}}} &= \mathbf{0}
\end{aligned}$$

In addition, the terms $\nabla_{\mathbf{x}} g(\mathbf{x}, \hat{\mathbf{n}}_j)$ and $\nabla_{\hat{\mathbf{n}}} g(\mathbf{x}, \hat{\mathbf{n}}_j)$ can be computed through chain rule, as follows

$$\begin{aligned}
g_0(\mathbf{x}, \hat{\mathbf{n}}) &= \frac{\langle \hat{\omega}_l(\mathbf{x}), \hat{\mathbf{n}} \rangle^2 \langle \hat{\omega}_l(\mathbf{x}), \hat{\mathbf{n}}(\mathbf{l}) \rangle^2}{\|\mathbf{x} - \mathbf{l}\|^4} \\
\frac{\partial g_0(\mathbf{x}, \hat{\mathbf{n}})}{\partial \mathbf{x}} &= \frac{2 \langle \hat{\omega}_l(\mathbf{x}), \hat{\mathbf{n}}(\mathbf{l}) \rangle \langle \hat{\omega}_l(\mathbf{x}), \hat{\mathbf{n}} \rangle}{\|\mathbf{x} - \mathbf{l}\|^5} \\
&\quad \cdot [\hat{\mathbf{n}}(\mathbf{l}) \langle \hat{\omega}_l(\mathbf{x}), \hat{\mathbf{n}} \rangle + \hat{\mathbf{n}} \langle \hat{\omega}_l(\mathbf{x}), \hat{\mathbf{n}}(\mathbf{l}) \rangle \\
&\quad + 4 \langle \hat{\omega}_l(\mathbf{x}), \hat{\mathbf{n}}(\mathbf{l}) \rangle \langle \hat{\omega}_l(\mathbf{x}), \hat{\mathbf{n}} \rangle \hat{\omega}_l(\mathbf{x})] \\
\frac{\partial g_0(\mathbf{x}, \hat{\mathbf{n}})}{\partial \hat{\mathbf{n}}} &= \frac{2 \langle \hat{\omega}_l(\mathbf{x}), \hat{\mathbf{n}}(\mathbf{l}) \rangle^2 \langle \hat{\omega}_l(\mathbf{x}), \hat{\mathbf{n}} \rangle}{\|\mathbf{x} - \mathbf{l}\|^4} \hat{\omega}_l(\mathbf{x})
\end{aligned}$$

Finally, the reflectance function is also a function of \mathbf{x} and $\hat{\mathbf{n}}$. Notice that ω_i and ω_o are functions of \mathbf{x} . We use symbolic differentiation to compute the derivative with respect to \mathbf{x} and $\hat{\mathbf{n}}$, respectively.

Accounting for visibility changes. Delaunoy and Prados [23] derived the gradient of a function that contains visibility terms for the case of surfaces represented as triangular meshes. In particular, they show how to differentiate energy functionals of the form,

$$E(\mathcal{S}) = \int_{\mathcal{S}} \langle \mathbf{g}(\mathbf{x}, \hat{\mathbf{n}}(\mathbf{x})), \hat{\mathbf{n}}(\mathbf{x}) \rangle V(\mathbf{x}, \mathbf{s}_0) dA(\mathbf{x}), \quad (5.12)$$

where \mathbf{s}_0 is the camera center for the line-of-sight imaging case. The visibility function checks whether a point is visible to the camera. Their derivative expression can be directly adapted to our setting in the case of confocal imaging $\mathbf{l} = \mathbf{s}$. In that case, we observe that the image formation model of Equation (5.1) simplifies to

$$I(t; \mathbf{l}) = \int_{\mathcal{S}_{\text{NLOS}}} g(\mathbf{x}, \hat{\mathbf{n}}(\mathbf{x})) V(\mathbf{x}, \mathbf{l}) dA(\mathbf{x}), \quad (5.13)$$

where

$$g(\mathbf{x}, \hat{\mathbf{n}}(\mathbf{x})) = f_s(\hat{\mathbf{n}}(\mathbf{x}), \hat{\boldsymbol{\omega}}_l(\mathbf{x}), \hat{\boldsymbol{\omega}}_l(\mathbf{x})) \cdot \frac{(\langle -\hat{\boldsymbol{\omega}}_l(\mathbf{x}), \hat{\mathbf{n}}(\mathbf{l}) \rangle)^2 (\langle \hat{\boldsymbol{\omega}}_l(\mathbf{x}), \hat{\mathbf{n}}(\mathbf{x}) \rangle)^2}{\|\mathbf{x} - \mathbf{l}\|^4}.$$

If we define

$$\mathbf{g}(\mathbf{x}, \hat{\mathbf{n}}(\mathbf{x})) = f_s(\hat{\mathbf{n}}(\mathbf{x}), \hat{\boldsymbol{\omega}}_l(\mathbf{x}), \hat{\boldsymbol{\omega}}_l(\mathbf{x})) \cdot \frac{(\langle -\hat{\boldsymbol{\omega}}_l(\mathbf{x}), \hat{\mathbf{n}}(\mathbf{l}) \rangle)^2 \langle \hat{\boldsymbol{\omega}}_l(\mathbf{x}), \hat{\mathbf{n}}(\mathbf{x}) \rangle}{\|\mathbf{x} - \mathbf{l}\|^4} \cdot \hat{\boldsymbol{\omega}}_l(\mathbf{x}),$$

then we can rewrite Equation (5.13) as

$$I(t; \mathbf{l}) = \int_{\mathcal{S}_{\text{NLOS}}} \langle \mathbf{g}(\mathbf{x}, \hat{\mathbf{n}}(\mathbf{x})), \hat{\mathbf{n}}(\mathbf{x}) \rangle V(\mathbf{x}, \mathbf{l}) dA(\mathbf{x}),$$

which now matches the form of Equation (5.12).

We describe the derivative expression by Delaunoy and Prados [23] for functionals of the form of Equation (5.12). The derivative contains three terms, terms due to the variation of the normal, term due to the tetrahedra of the visible adjacent triangles, and the term due to the movement of the crepuscular cone. Please refer to [23] for the details.

$$\begin{aligned} \frac{\partial E(\mathcal{S})}{\partial \mathbf{v}_i} &= \mathbf{G}_i^{\text{norm}} + \mathbf{G}_i^{\text{int}} + \mathbf{G}_i^{\text{horiz}} \\ \mathbf{G}_i^{\text{norm}} &= - \sum_{k \in J_i} \frac{\mathbf{e}_{k,i}}{2A_k} \wedge \int_{\mathcal{T}_k} P_{\hat{\mathbf{n}}_k} (D_{\hat{\mathbf{n}}_k} \mathbf{g}(\mathbf{x}, \mathbf{n}_k)^T \hat{\mathbf{n}}_k) V(\mathbf{x}, \mathbf{s}_0) ds \\ \mathbf{G}_i^{\text{int}} &= \sum_{k \in J_i} \hat{\mathbf{n}}_k \int_{\mathcal{T}_k} \nabla \cdot \mathbf{g}(\mathbf{x}, \hat{\mathbf{n}}_k) \phi_i(\mathbf{x}) ds \\ \mathbf{G}_i^{\text{horiz}} &= \sum_{\mathbf{H}_{i,k}} \frac{1}{2} \int_0^1 \{ [p(T(\mathbf{y}(u))) - p(\mathbf{y}(u))] \\ &\quad \left[\frac{\mathbf{y}(u)}{\|\mathbf{y}(u)\|^4} \wedge \mathbf{H}_{i,k} \right] (1-u) \} du \end{aligned}$$

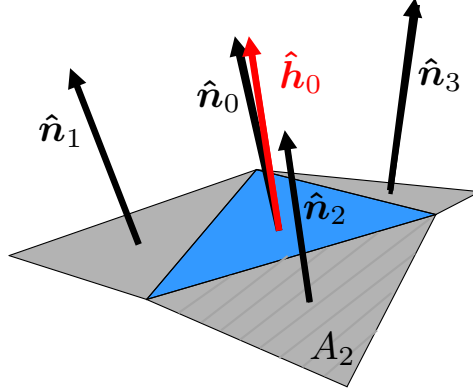


Figure 5.2: **Normal smoothness.** For surface smoothness regularization, we optimize vertices around a face so that the normal is aligned with the weighted normal around the face.

$P_{\hat{\mathbf{n}}_k}(\cdot)$ is the projection on the orthogonal plane to $\hat{\mathbf{n}}_k$. A_k is the area of face k . $\mathbf{H}_{i,k}$ is the vector such that $[\mathbf{v}_i, \mathbf{v}_i + \mathbf{H}_{i,k}]$ is the edge of triangle k generating the horizon. \mathbf{y} corresponds to points sampled on the edge $\mathbf{H}_{i,k}$. $T(x)$ corresponds to the point located behind x in the direction of the viewpoint. $\mathbf{g}(x) = p(x) \frac{x}{x_z^2}$

Surface regularization. We note that, when optimizing geometry, we follow Delaunoy and Prados [23] and augment the loss function $E(\mathbf{v}, \boldsymbol{\pi})$ with a normal smoothing regularization term. We can define the regularization energy as

$$\mathcal{R}(\mathbf{v}) = \sum_k A_k \left(1 - \hat{\mathbf{h}}_k \cdot \hat{\mathbf{n}}_k\right),$$

where A_k and $\hat{\mathbf{n}}_k$ are the area and face normal, respectively, of the k -th mesh triangle, and $\hat{\mathbf{h}}_k$ is the weighted average of the face normals of all triangles in its neighborhood N_k ,

$$\hat{\mathbf{h}}_k = \frac{\sum_{i \in N_k} A_i \hat{\mathbf{n}}_i}{\left\| \sum_{i \in N_k} A_i \hat{\mathbf{n}}_i \right\|}.$$

Its derivative with respect to mesh vertices is provided by Delaunoy and Prados [23].

5.3.2 Stochastic estimation and optimization

We now describe our computational tools for efficiently minimizing the loss function of Equation (5.4). We show an overview of our optimization pipeline in Figure 5.1 and Algorithm 6, and discuss below the various components and strategies we use to make convergence faster and more robust.

Monte Carlo rendering. The surface integrals of Equations (5.1), (5.11), (5.10) can be approximated using Monte Carlo integration: We first use any probability distribution μ on $\mathcal{S}_{\text{NLOS}}$ to

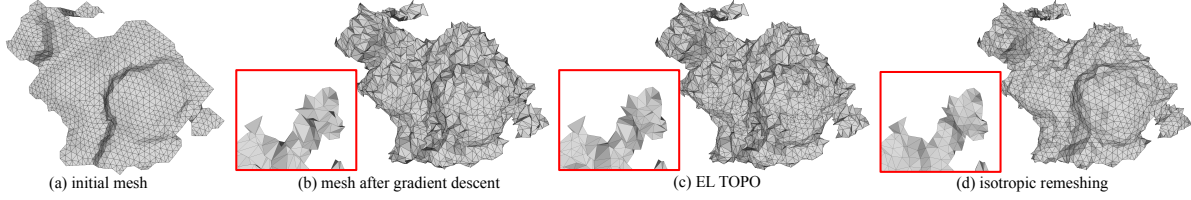


Figure 5.3: **Geometry processing.** We show an example of the effects of the geometric processing operations we use during mesh optimization to improve mesh topology. (a) Initial mesh. (b) Mesh after gradient descent steps. As shown in the inset, at several places the mesh is self-intersecting. (c) Mesh evolution using El Topo, which helps reduce self-intersections. (d) Mesh after isotropic remeshing, which increases mesh detail while decreasing high-frequency artifacts.

sample a set of points $\{\mathbf{x}_j \in \mathcal{S}_{\text{NLOS}}, j = 1, \dots, J\}$. Then, we can form the respective unbiased and consistent estimates [30]:

$$\begin{aligned} \langle I \rangle &= \sum_{j=1}^J \frac{g(\mathbf{x}_j, \hat{\mathbf{n}}(\mathbf{x}_j)) V(\mathbf{x}_j, \mathbf{l}) V(\mathbf{x}_j, \mathbf{s})}{\mu(\mathbf{x}_j)}, \\ \left\langle \frac{\partial I}{\partial \pi} \right\rangle &= \sum_{j=1}^J \frac{g_r(\mathbf{x}_j, \hat{\mathbf{n}}(\mathbf{x}_j)) V(\mathbf{x}_j, \mathbf{l}) V(\mathbf{x}_j, \mathbf{s})}{\mu(\mathbf{x}_j)}, \\ \left\langle \frac{\partial I}{\partial v} \right\rangle &= \sum_{j=1}^J \frac{g_s(\mathbf{x}_j, \hat{\mathbf{n}}(\mathbf{x}_j)) V(\mathbf{x}_j, \mathbf{l}) V(\mathbf{x}_j, \mathbf{s})}{\mu(\mathbf{x}_j)}. \end{aligned} \quad (5.14)$$

Next, we describe a stratified area sampling procedure, which greatly accelerates rendering.

Monte Carlo rendering with stratified area sampling. We first discuss the rendering algorithm we use to estimate transients $I(t; \mathbf{l}, \mathbf{s})$, and their derivatives with respect to the surface, $\frac{\partial I}{\partial v}$, and with respect to reflectance, $\frac{\partial I}{\partial \pi}$ (Equations (5.1), (5.10), and (5.11), respectively). We present the algorithm in the context of estimating $I(t; \mathbf{l}, \mathbf{s})$, but the discussion applies exactly for the $\frac{\partial I}{\partial v}$ and $\frac{\partial I}{\partial \pi}$ cases. Algorithm 7 shows an overview of our rendering procedure.

We use Monte Carlo integration to approximate the surface integral of Equation (5.1): We first use any probability distribution μ on $\mathcal{S}_{\text{NLOS}}$ to sample a set of points $\{\mathbf{x}_j \in \mathcal{S}_{\text{NLOS}}, j = 1, \dots, J\}$. Then, we form an estimate as in Equation (5.14). In the context of light transport simulation, this is referred to as Monte Carlo rendering [27, 106, 131]. We note that Monte Carlo integration by sampling points \mathbf{x} on $\mathcal{S}_{\text{NLOS}}$ is equivalent to the area sampling strategy described by Veach and Guibas [133]. The fact that we use area sampling is also the reason for the presence of the shading terms $\langle -\hat{\omega}_l(\mathbf{x}), \hat{\mathbf{n}}(\mathbf{l}) \rangle \langle \hat{\omega}_l(\mathbf{x}), \hat{\mathbf{n}}(\mathbf{x}) \rangle / \|\mathbf{x} - \mathbf{l}\|^2$ and $\langle -\hat{\omega}_s(\mathbf{x}), \hat{\mathbf{n}}(\mathbf{s}) \rangle \langle \hat{\omega}_s(\mathbf{x}), \hat{\mathbf{n}}(\mathbf{x}) \rangle / \|\mathbf{x} - \mathbf{s}\|^2$ in the estimate of Equation (5.14) (see Equation (9) in Veach and Guibas [133]).

We note that standard path sampling algorithms (e.g., path tracing [62], bidirectional path tracing [132], Metropolis light transport [134]) typically use directional sampling instead of area sampling. However, area sampling is advantageous in our setting for two reasons: First, when

the NLOS surface $\mathcal{S}_{\text{NLOS}}$ is small compared to the LOS surface \mathcal{S}_{LOS} , directional sampling will result in a large number of missed rays, greatly reducing rendering efficiency. This has also been observed by Klein et al. [70], who report an order of magnitude efficiency improvement when using area sampling instead of directional sampling.

Second, area sampling facilitates stratified sampling [30]. We start by noting that, when $\mathcal{S}_{\text{NLOS}}$ is a triangular mesh, Equation (5.1) can be decomposed into a sum of per-triangle integrals,

$$I = \sum_{k=1}^K \int_{\mathcal{T}_k} g(\mathbf{x}, \hat{\mathbf{n}}(\mathbf{x})) V(\mathbf{x}, \mathbf{l}) V(\mathbf{x}, \mathbf{s}) \, dA(\mathbf{x}), \quad (5.15)$$

where \mathcal{T}_k is the k -th mesh triangle, and $\mathcal{S}_{\text{NLOS}} = \cup_{k=1}^K \mathcal{T}_k$. We can then estimate I as

$$\langle I \rangle = \sum_{k=1}^K \sum_{j=1}^{\lceil J/K \rceil} \frac{g(\mathbf{x}_j^k, \hat{\mathbf{n}}(\mathbf{x}_j^k)) V(\mathbf{x}_j^k, \mathbf{l}) V(\mathbf{x}_j^k, \mathbf{s})}{\mu(\mathbf{x}_j^k)}, \quad (5.16)$$

where for each k , $\{\mathbf{x}_j^k \in \mathcal{T}_k, j = 1, \dots, \lceil J/K \rceil\}$. When μ is the uniform distribution, Equation (5.14) uses J points \mathbf{x}_j uniformly sampled from the entire mesh $\mathcal{S}_{\text{NLOS}}$. By contrast, Equation (5.16) splits the J points into K equal-sized sets, with points in the k -th set uniformly sampled from a stratum corresponding to the triangle \mathcal{T}_k . This stratification procedure can significantly reduce the variance of $\langle I \rangle$ [88, 108], by up to a factor of $1/K$ if the integrand of Equation (5.15) is approximately constant within each triangle—as is the case for meshes with very fine triangulation. Empirically, we observed that stratified sampling reduces variance by an order of magnitude compared to uniform sampling.

An additional advantage of stratified sampling becomes evident when we consider the visibility terms V in Equations (5.14) and (5.16). For large meshes, visibility tests can account for the bulk of the rendering computational cost. Therefore, it is critical that these tests be performed using highly-optimized libraries specifically designed for this task [98, 138]. When using stratified sampling, the ray bundles $\{s \rightarrow \mathbf{x}_j^k\}$ and $\{\mathbf{l} \rightarrow \mathbf{x}_j^k\}$ for all points of the same k can be treated as coherent ray packets for the purposes of visibility testing [107, 137]. Empirically, we found that this results in a fourfold rendering acceleration, highlighting another advantage of our triangle-based stratification procedure.

We conclude this discussion by noting one important disadvantage of our area sampling procedure (either stratified or otherwise): When the reflectance f_s of the underlying surface becomes highly specular, then the estimate of Equation (5.14) becomes very inefficient. This is because, when \mathbf{x}_j is sampled uniformly on $\mathcal{S}_{\text{NLOS}}$ or a triangle \mathcal{T}_k , the half-vector corresponding to the path $\mathbf{l} \rightarrow \mathbf{x}_j \rightarrow \mathbf{s}$ will deviate significantly from the normal $\hat{\mathbf{n}}(\mathbf{x}_j)$ with high probability. As a result, most of the sampled paths will have very low integrand values $g(\mathbf{x}_j, \hat{\mathbf{n}}(\mathbf{x}_j))$. The effect of this can be observed in the experiment section, Figure 5.10, when optimizing for the NLOS case under very specular BRDF (GGX with $\alpha = 0.1$). This can be overcome by incorporating into our framework area sampling techniques that can account for BRDF effects through half-vector importance sampling [64] and we leave it for future work.

Stochastic gradient descent. Using these Monte Carlo estimates, we can approximately compute the derivatives of Equation (5.5). We can combine these stochastic derivative estimates with stochastic gradient descent (SGD) algorithms to perform the analysis-by-synthesis optimization of Equation (5.4). In particular, we use the Adam [68] variant of SGD, in order to take advantage of the adaptive per-parameter learning-rate scheduling, to alternately optimize for reflectance and surface.

When optimizing for surface vertices, Adam ordinarily would maintain $3 \cdot V$ independent learning rates, where V is the number of vertices, and each vertex is a three-dimensional vector. Empirically, we found it beneficial to constrain all 3 coordinates of the same vertex to have the same learning rate, reducing the independent learning rates to V —one per 3D vertex vector. To achieve this, we modified Adam to update the learning rate for each vertex using the total magnitude of the 3D gradient for that vertex’s coordinates. For the learning rate update rules, we refer to the original Adam publication [68].

Continuation. Following common practice in regularized optimization [46], we adopt a continuation scheme and gradually decrease the weight λ of the regularization term. We start with an initial value λ_0 , and decrease λ by a factor of 1.25 every time we increase the surface resolution. This gives more and more emphasis on the data term instead of the regularization.

$$E(v, \pi) + \lambda \mathcal{R}(v).$$

Rendering sample budgeting. We follow the so-called increasing precision strategy proposed by Pfeiffer and Sato [105]. As the optimization proceeds, we increase the number of samples J used for rendering the gradients with respect to surface and reflectance. Intuitively, as the optimization gets closer to a local minimum, and therefore the true gradient becomes smaller, the gradient-descent procedure becomes more sensitive to variance in the gradient estimates; therefore, reducing this variance by increasing the rendering samples can help convergence. In our implementation, we use an initial number of J_0 samples, and we increase this by a factor of 1.25 every time we alternate between the surface and reflectance optimization.

Comparison with numerical differentiation. The performance of SGD optimization critically depends on the ability to compute unbiased gradient estimates of low variance. To highlight the importance of our differentiable rendering formulation in facilitating this optimization, we compare rendered gradient estimates with estimates computed using a finite-difference approximation in Figure 5.4. We observe that the numerical gradients have significantly higher variance; therefore, using them with SGD would greatly slow down convergence. Alternatively, we could eliminate variance in finite-difference estimation, by using a quadrature technique (e.g., finite elements) to compute the forward integral (5.1). However, this could introduce strong bias, and therefore affect the physical accuracy of the result.

5.4 Geometry processing operations

As discussed in Section 5.3.1, during our analysis-by-synthesis procedure, we use stochastic gradient descent to only optimize for $\mathcal{S}_{\text{NLOS}}$ mesh vertices and not topology. We describe our use

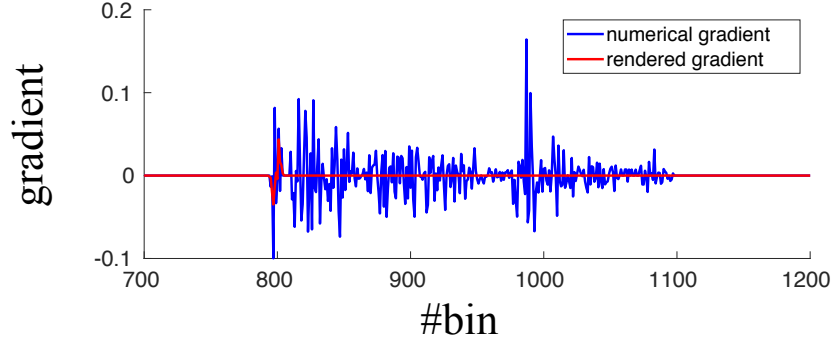


Figure 5.4: **Comparison of gradient estimation techniques.** We estimate the derivative of a transient with respect to one coordinate of one NLOS surface vertex. We plot two estimates, one computed using the rendering algorithm of Section 5.3.2, and another using finite differences and antithetic variates [30]. We observe that the numerical gradient is significantly noisier than the rendered one, despite being computed using twice as many path samples.

of geometry processing tools, at various times during the analysis-by-synthesis optimization, to improve the topology of the mesh $\mathcal{S}_{\text{NLOS}}$. Figure 5.3 shows example instances of these geometry processing operations.

Robust surface evolution. As the mesh $\mathcal{S}_{\text{NLOS}}$ evolves over multiple SGD iterations, triangle quality typically degrades, and self-intersections occur. Motivated from other optimization-driven surface editing algorithms [79], we instead evolve $\mathcal{S}_{\text{NLOS}}$ using the pipeline proposed by Brochu and Bridson [14], implemented in El Topo [13]. Given an initial set of vertices $\{v_i, i = 1, \dots, V\}$, and associated displacement vectors $\{\mathbf{d}v_i, i = 1, \dots, V\}$, El Topo performs two types of operations: First, it alters the displacement vectors and mesh topology, to produce a non-intersecting mesh. Second, it performs local topology operations to improve overall mesh quality. In our implementation, we accumulate displacement vectors $\mathbf{d}v_i$ over multiple gradient-descent iterations, then use El Topo to evolve the mesh.

Progressive refinement and isotropic remeshing. As an additional means of regularization and acceleration, we optimize the NLOS surface $\mathcal{S}_{\text{NLOS}}$ in a coarse-to-fine fashion. We start with a mesh of a relatively small number of vertices V and triangles T . Then, during the gradient-descent optimization of $\mathcal{S}_{\text{NLOS}}$, we progressively increase the number of vertices and triangles. We implement mesh refinement by first use El Topo [13] to create a non-intersecting version of the evolved mesh, then perform isotropic remeshing [11] with an increased number of target vertices.

5.5 Experiments

Implementation. Our NLOS reconstruction framework has three major components, differentiable rendering, geometry processing, and stochastic gradient descent. For rendering, we have developed two C++ implementations: The first targets CPU execution and uses Embree [138] to accelerate rendering. The second is GPU-based and relies on OptiX [98] for fast rendering. For

geometry processing, our C++ implementation is built using the El Topo [13], CGAL [123] and libigl [97] libraries. Finally, both the rendering and geometry processing components are interfaced with Pytorch [99], which we use for stochastic gradient descent optimization with the Adam algorithm [68]. Our implementation can scale up to surface optimization problems on meshes with more than 100,000 vertices, using 4096 transient measurements of 1200 temporal bins each. We will make our implementation publicly available. We will additionally release our synthetic and real datasets.

Scanning configuration. In all of our experiments, we use the confocal scanning procedure proposed by O’Toole et al. [96], where measurements are captured by illuminating and sensing at the same point on the visible surface ($\mathbf{l} = \mathbf{s}$). The scanning points on the wall are on a regular grid of 64×64 points, at different spatial resolutions.

Initialization. We initialize our optimization using the volumetric reconstruction produced by the method of O’Toole et al. [96], using the implementation provided by the authors. We convert the volumetric representation into a surface by first computing the maximum albedo voxel along the depth axis, pruning voxels with albedo below some threshold, and finally triangulating the remaining points.

5.5.1 Simulation results

We use synthetic data to evaluate the ability of our method to reconstruct NLOS surface shape and reflectance. In our synthetic experiments, NLOS objects are placed at a distance of 0.4 m from a visible wall of size $0.5m \times 0.5m$. We use physically-accurate Monte Carlo rendering to synthesize data, and use the model of [51] to add noise.

Shape reconstruction. Figure 5.5 shows reconstructions for a variety of NLOS shapes with a known Lambertian reflectance. We observe that our method can reconstruct surface details that are completely missing from the volumetric reconstruction used for initialization. A notable result is the soap, where we can reconstruct the relief letters (depth 2 transient bins).

Different initialization method. The energy function (5.4) we seek to optimize is non-convex. Therefore, the performance is heavily dependent on the initialization. We test the optimization framework by initializing it with solution of space carving (Algorithm 3). As shown in Figure 5.6, we can see that our proposed method can improve the surface estimation from a low resolution and low surface detail spatial extend representation to something closer to ground truth. For the simpler case in the last column, the soap scene, the optimized surface reveals the text relief. However, as seen in the first column, the armadillo scene consist of consists of minute details and non-convex surface geometry and hence initialization with space carving is significantly far away from ground truth, resulting in a solution that converges further away.

Number of measurements. We compare reconstructions obtained by the technique of O’Toole et al. [96] and our procedure, using measurements from confocal scans ($\mathbf{l} = \mathbf{s}$) at 5 different scanning resolutions within the same scanning area on the visible wall \mathcal{S}_{LOS} . In particular, we consider resolutions: 16×16 , 32×32 , 64×64 , 128×128 , and 256×256 scan points. Figure 5.7 shows the results. We note that, for the 128×128 and 256×256 cases, we found it beneficial to initialize

our optimization procedure using a reconstruction obtained by applying the algorithm of O’Toole et al. [96] on only 64×64 measurements. Despite the loss in detail, this initialization provided better coverage on the NLOS surface.

We observe that, even in the case of 16×16 measured transients, our algorithm can recover some of the detail on the surface of the ground-truth mesh (e.g., the texture on the bunny’s leg), despite the fact that the initialization has no visible details. We also see that, with just 32×32 measured transients, our algorithm can reconstruct more detail than what is possible using the volumetric reconstruction from O’Toole et al. [96] with 256×256 measured transients. That is, we can obtain a more detailed reconstruction even though we are using 64 times fewer measurements.

Noise levels. In this experiment, we evaluate the performance of our method as a function of the amount of noise contaminating our transient measurements. We use synthetic transients, and simulate SPAD noise using the method of Hernandez et al. [51]. The simulated noise contains three components, ambient noise, Poisson noise, and SPAD jitter.

Figure 5.8 shows reconstructions under three different noise levels, corresponding to different number of laser pulses M . Even though performance deteriorates as noise increases, we see that our proposed method can take SPAD jitter into account and is robust to ambient and Poisson noise.

We note that, as the noise level increases, the level of detail we can recover decreases. In this case, continuing to increase the spatial resolution of the mesh can be counter-productive, as the amount of noise means that gradient estimates have very high variance: We simply have not measured enough photons in order for our measurements to sufficiently regularize the surface $\mathcal{S}_{\text{NLOS}}$. This is shown in Figure 5.9, where we show the evolution of the mesh as we continue to upsample it at higher spatial resolutions. In practice, this implies that, for higher noise levels, we need to stop the optimization procedure at earlier resolution levels.

Simultaneous shape and reflectance reconstruction. Figure 5.10 shows simulated experiments for reconstructing both shape and reflectance. We experiment with a range of GGX α values, going from very smooth to very rough specular reflectance (Figure 5.10(b)-(c)). We observe that our algorithm successfully reconstructs a rough estimate of both shape and reflectance in all cases, but the reconstruction quality deteriorates as the surface becomes more specular.

5.5.2 Real scene

We additionally perform experiments using three datasets captured with SPAD-based transient imaging systems for real NLOS scenes. The first dataset is the diffuse ‘S’ shape object from [96]. As shown in Figure 5.11, our recovered shape closely resembles the ground truth and is overall flatter, matching the ground truth geometry.

We additionally show reconstructions for two datasets captured with our own implementation of the SPAD setup of [96], for two NLOS objects of greater surface complexity. The first object is a diffuse horse bust with fine geometric details. As shown in Figure 5.11, our recovered result reproduces the flat and curved surface areas better. The second object is a planar scene with a 6mm tall (5 transient bins) relief in the shape of two digits. Our recovered result better differentiates the digits from the background surface.

5.6 Discussion and conclusion

In this chapter, we introduce an analysis-by-synthesis framework that can reconstruct detailed shape and complex reflectance of an NLOS object. Our framework deviates from prior work on NLOS reconstruction, by directly optimizing for a surface representation of the NLOS object, in place of commonly employed volumetric representation. Additionally, this optimization is performed while accurately modeling the underlying light transport physics through the rendering equation. At the core of our framework is a new rendering formulation that can be used to efficiently compute derivatives of radiometric measurements, such as transients, with respect to NLOS geometry and reflectance. By coupling this with stochastic optimization and geometry processing techniques, we are able to reconstruct NLOS surface at a level of detail significantly exceeding what is possible with previous volumetric reconstruction methods.

In the following, we discuss some improvements of our NLOS surface optimization framework that can be incorporated in the future.

Dependency on initialization. We pose our reconstruction as an optimization problem with a very non-linear energy function, therefore, our final reconstruction is strongly dependent on a good initialization. Our experiments indicate that, if the initialization misses a large part of the NLOS object, our optimization process will have difficulty recovering that part. We hope to address this by exploring alternate initialization schemes, and by incorporating boundary evolution techniques into our optimization pipeline.

Handling specular reflectance. The performance of our pipeline degrades as the reflectance of the NLOS object becomes more specular. We believe this is primarily caused by the area sampling procedure we use for rendering, which becomes very inefficient for highly-specular reflectance. We can potentially improve performance in such cases by considering importance sampling techniques for rendering.

Noise model. The energy function (5.4) we seek to minimize is equivalent to model Gaussian noise. However, since we use a SPAD sensor to capture the light transient, using Poisson distribution to model noise will be more accurate.

In the following, we show relevant modifications to the optimization process that models Poisson process. We optimize the surface and reflectance parameters such that negative log likelihood is minimized.

$$\begin{aligned}
& \min_{\mathbf{v}, \boldsymbol{\pi}} -\log P(\tilde{I} | I[\mathbf{v}, \boldsymbol{\pi}]) \\
& \equiv \min_{\mathbf{v}, \boldsymbol{\pi}} -\log \prod_{m,t} \left(e^{-I[\mathbf{v}, \boldsymbol{\pi}](t; \mathbf{l}_m, \mathbf{s}_m)} \frac{I[\mathbf{v}, \boldsymbol{\pi}](t; \mathbf{l}_m, \mathbf{s}_m)^{\tilde{I}_m(t)}}{\tilde{I}_m(t)!} \right) \\
& \equiv \min_{\mathbf{v}, \boldsymbol{\pi}} -\sum_{m,t} -I[\mathbf{v}, \boldsymbol{\pi}](t; \mathbf{l}_m, \mathbf{s}_m) + \tilde{I}_m(t) \log[I[\mathbf{v}, \boldsymbol{\pi}](t; \mathbf{l}_m, \mathbf{s}_m)] - \log[\tilde{I}_m(t)!] \\
& \equiv \min_{\mathbf{v}, \boldsymbol{\pi}} \sum_{m,t} I[\mathbf{v}, \boldsymbol{\pi}](t; \mathbf{l}_m, \mathbf{s}_m) - \tilde{I}_m(t) \log[I[\mathbf{v}, \boldsymbol{\pi}](t; \mathbf{l}_m, \mathbf{s}_m)]
\end{aligned}$$

We define the new energy function as

$$E_{\text{Pois}}(\boldsymbol{v}, \boldsymbol{\pi}) = \sum_{m,t} I[\boldsymbol{v}, \boldsymbol{\pi}](t; \boldsymbol{l}_m, \boldsymbol{s}_m) - \tilde{I}_m(t) \log[I[\boldsymbol{v}, \boldsymbol{\pi}](t; \boldsymbol{l}_m, \boldsymbol{s}_m)] \quad (5.17)$$

Same as how we optimize (5.4), we take the derivative of (5.17) with respect to \boldsymbol{v} or $\boldsymbol{\pi}$.

$$\frac{\partial E_{\text{Pois}}(\boldsymbol{v}, \boldsymbol{\pi})}{\partial \boldsymbol{y}} = - \sum_{m,t} \left[\frac{\tilde{I}_m(t)}{I[\boldsymbol{v}, \boldsymbol{\pi}](t; \boldsymbol{l}_m, \boldsymbol{s}_m)} - 1 \right] \frac{\partial I[\boldsymbol{v}, \boldsymbol{\pi}](t; \boldsymbol{l}_m, \boldsymbol{s}_m)}{\partial \boldsymbol{y}}$$

where \boldsymbol{y} can be either \boldsymbol{v} or $\boldsymbol{\pi}$.

Applicable to all types of radiometric measurements Our framework can be used to process not only transients, but all types of radiometric measurements: steady-state, continuous-wave time-of-flight, and so on, since the framework is based on the rendering equation. Therefore, we hope it can serve as a platform for exploring NLOS scanning schemes that use alternative radiometric sensors, or even fuse together measurements from multiple types of sensors.

Algorithm 6 Surface and reflectance optimization pipeline.

- Require:** Initial NLOS vertices v_0 .
Require: Initial number of NLOS vertices V_0 .
Require: Initial NLOS triangles T_0 .
Require: Initial reflectance parameters π_0 .
Require: Initial learning rates $\eta_{s,0}, \eta_{r,0}$.
Require: Measured transient \tilde{I} .
Require: Virtual source l and detector s .
Require: Initial regularization weight λ_0 .
Require: Initial number of rendering samples J_0 .
Require: Number of iterations N .
-

Initialization.

- 1: $v \leftarrow v_0$.
 - 2: $V \leftarrow V_0$.
 - 3: $T \leftarrow T_0$.
 - 4: $\pi \leftarrow \pi_0$.
 - 5: $\lambda \leftarrow \lambda_0$.
 - 6: $\eta_s \leftarrow \eta_{s,0}$.
 - 7: $\eta_r \leftarrow \eta_{r,0}$.
 - 8: $J \leftarrow J_0$.
 - 9: **while** not converged **do**
-

Gradient-descent optimization.

- 10: $\mathcal{S}_{\text{NLOS}} \leftarrow \text{CreateMesh}(v, T)$.
 - 11: ▷ Update reflectance.
 - 12: $\pi \leftarrow \text{OptReflectance}(\pi, \eta_r, \mathcal{S}_{\text{NLOS}}, \tilde{I}, l, s, N, J)$.
 - 13: ▷ Update surface.
 - 14: $v' \leftarrow \text{OptSurface}(v, \eta_s, \pi, T, \tilde{I}, l, s, \lambda, N, J)$.
-

Geometry processing.

- 15: ▷ Deform mesh.
 - 16: $(v, T) \leftarrow \text{ElTopo}(v, v', T)$.
 - 17: ▷ Increase mesh resolution.
 - 18: $V \leftarrow V \cdot 1.25$.
 - 19: $(v, T) \leftarrow \text{IsoRemesh}(v, T, V)$.
-

Update parameters.

- 20: $\lambda \leftarrow \lambda / 1.25$. ▷ Continuation.
 - 21: $J \leftarrow J \cdot 1.25$. ▷ Increasing precision.
 - 22: **end while**
-

Algorithm 7 Monte Carlo rendering for estimating transient I , or its derivatives $\frac{\partial I}{\partial v}$, $\frac{\partial I}{\partial \pi}$.

Require: Integrand function $G \in \{g, g_s, g_r\}$.

Require: Mesh $\mathcal{S}_{\text{NLOS}} = \bigcup_{k=1}^K \mathcal{T}_k$.

Require: Reflectance parameters π .

Require: Virtual source l and detector s .

Require: Number of rendering samples J .

```

1: function RENDER( $G, \mathcal{S}_{\text{NLOS}}, \pi, l, s, J$ )
2:    $\langle I \rangle \leftarrow 0$ . ▷ Initialize estimate.
3:   for  $k \in \{1, \dots, K\}$  do ▷ Select triangle  $\mathcal{T}_k$ .
4:   ▷ Compute triangle quantities.
5:      $A \leftarrow \text{ComputeTriangleArea}(\mathcal{T}_k)$ .
6:      $\hat{n} \leftarrow \text{ComputeTriangleNormal}(\mathcal{T}_k)$ .
7:     for  $j \in \{1, \dots, \lceil J/K \rceil\}$  do
8:       ▷ Uniformly sample a point on the triangle.
9:        $x \leftarrow \text{SamplePointUniformly}(\mathcal{T}_k)$ .
10:      ▷ Evaluate the function to be integrated.
11:       $y \leftarrow \text{EvaluateIntegrand}(G; \pi, x, \hat{n}, l, s)$ .
12:      ▷ Perform visibility tests.
13:       $v_l \leftarrow \text{EvaluateVisibility}(\mathcal{S}_{\text{NLOS}}, x, l)$ .
14:       $v_s \leftarrow \text{EvaluateVisibility}(\mathcal{S}_{\text{NLOS}}, x, s)$ .
15:       $\langle I \rangle \leftarrow \langle I \rangle + y \cdot v_l \cdot v_s \cdot A / \lceil J/K \rceil$ .
16:     end for
17:   end for
18:   return  $\langle I \rangle$ .
19: end function

```

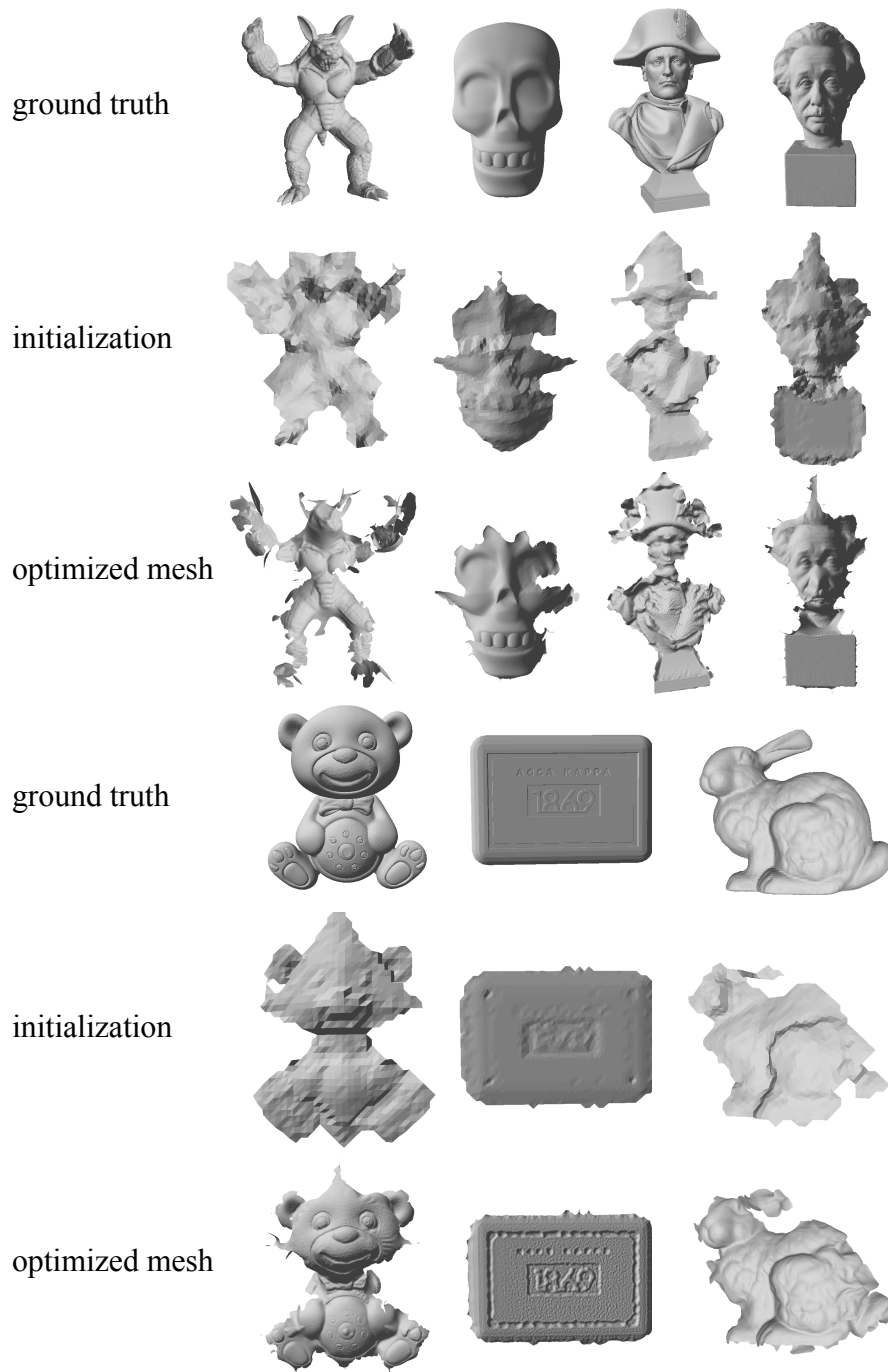


Figure 5.5: **NLOS surface reconstruction results of multiple types of surface detail.** Initialization results generated by codebase provided by O’Toole *et al.* [96]. Our procedure can reconstruct shapes with different surface characteristics, including highly non-convex shape, large depth variations, and bas-relief text.

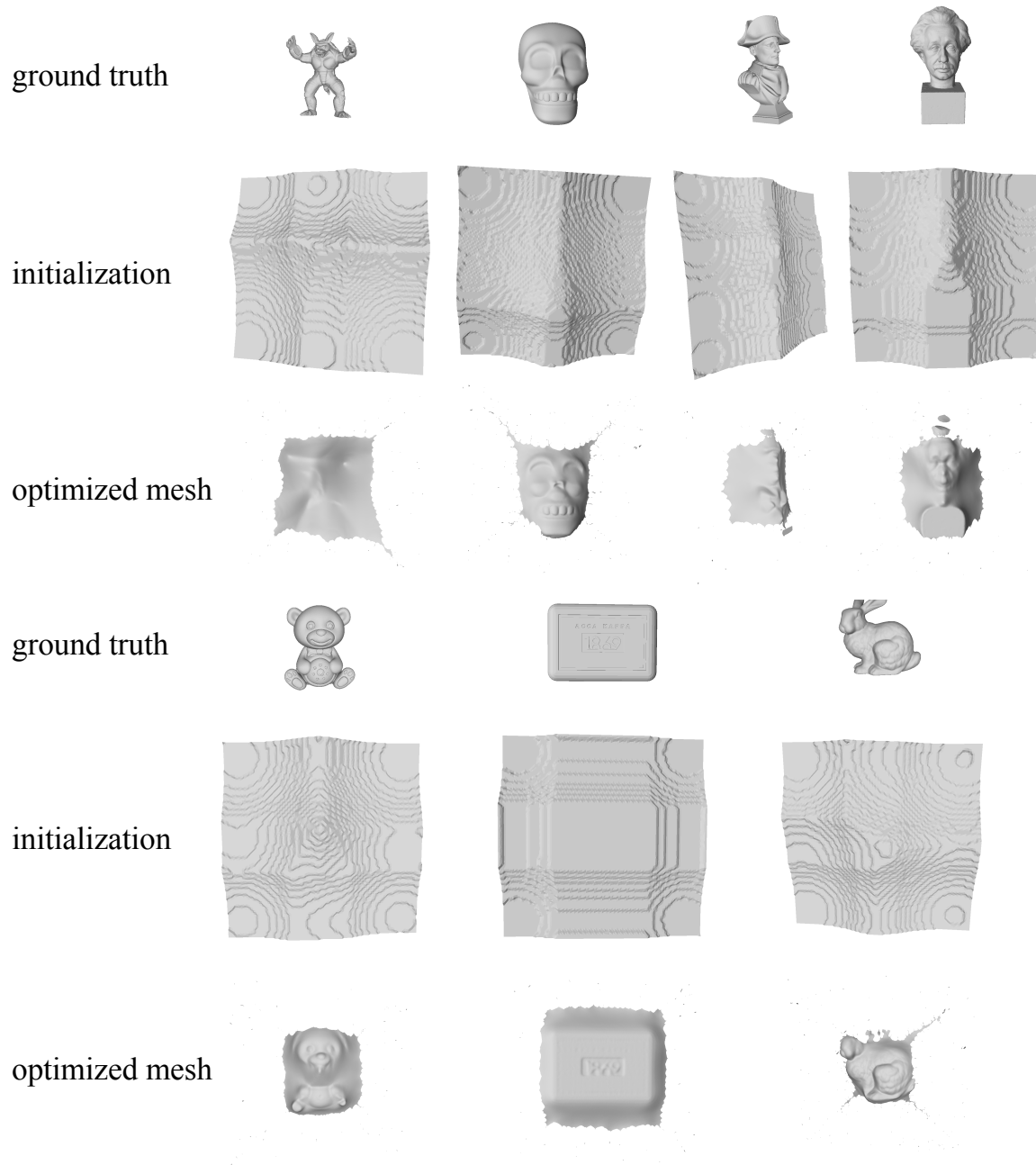


Figure 5.6: **NLOS surface reconstruction results with an alternative initialization method.** Initialization results generated by Algorithm 3. Our shape optimization framework can also work with different initialization method.

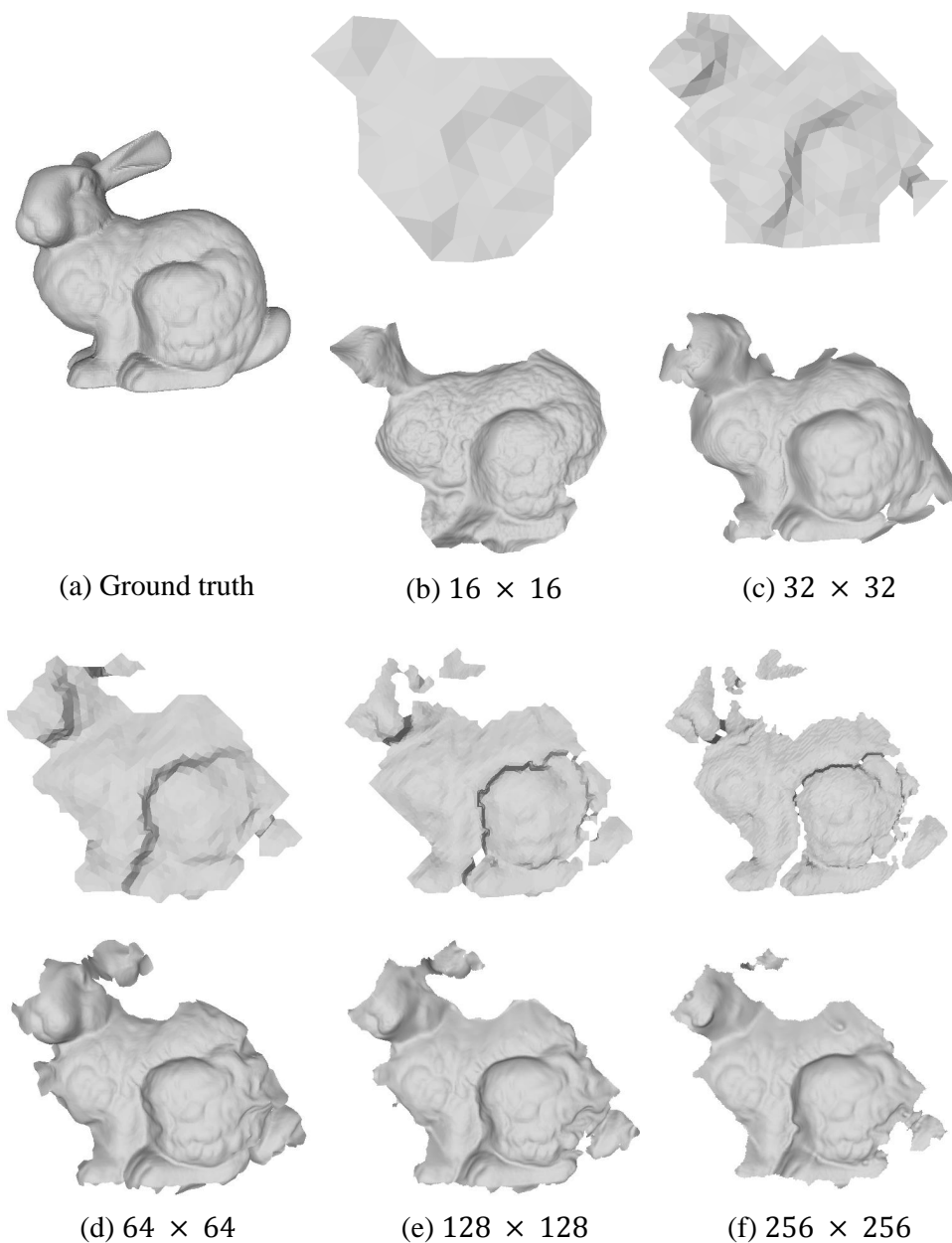


Figure 5.7: **Surface optimization using different numbers of measurements.** We perform experiments for different numbers of measured transients (i.e., scanned points on the LOS surface \mathcal{S}_{LOS}). When the number of measurements is very small, the initialization does not recover any discernible shape, whereas our surface optimization framework still recovers details of the ground-truth mesh. As the number of measurements increases, the level of detail of both the initialization and our reconstruction increases; in all cases, our method significantly improves the final surface reconstruction.

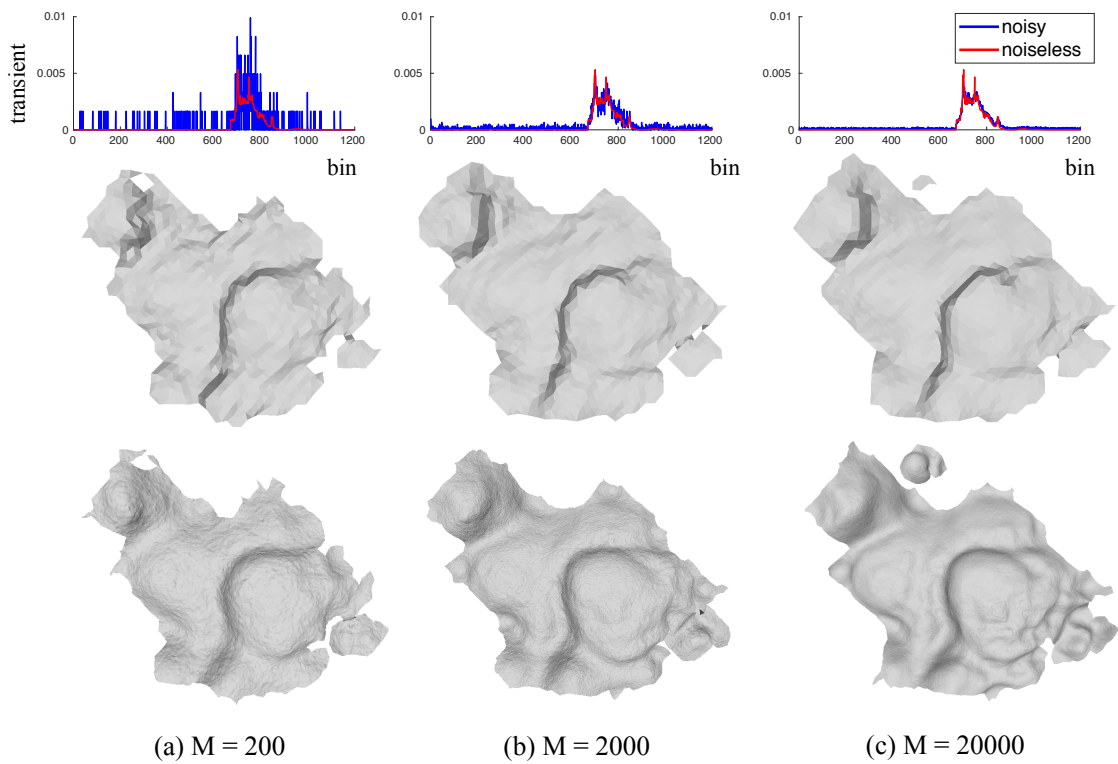


Figure 5.8: **NLOS surface optimization with simulated SPAD Noise** We simulate measurement noise by including ambient light, Poisson noise, and SPAD jitter. (Top) Sample of the noisy transients. (Middle) Initialization (Bottom) Our recovered results.

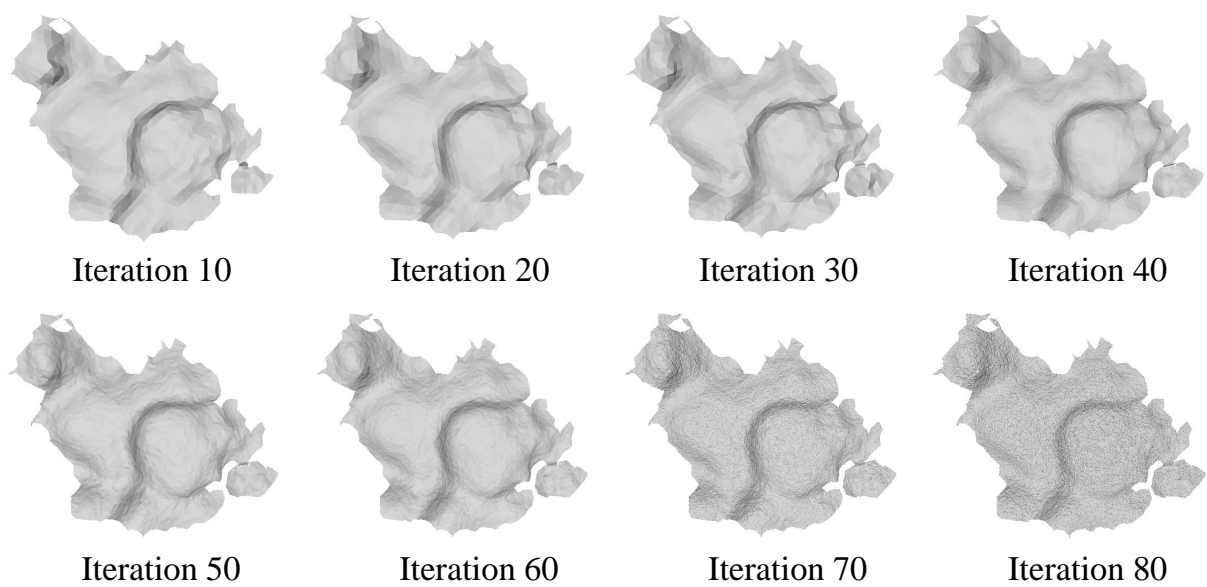


Figure 5.9: **Surface progression of noise experiment.** We show the progression of the surface optimization for the case of the highest noise ($M = 200$). The number of mesh vertices increases with iteration. As the spatial resolution of the surface increases, the effect of noise is to reduce local surface quality, even if the global surface shape remains acceptable.

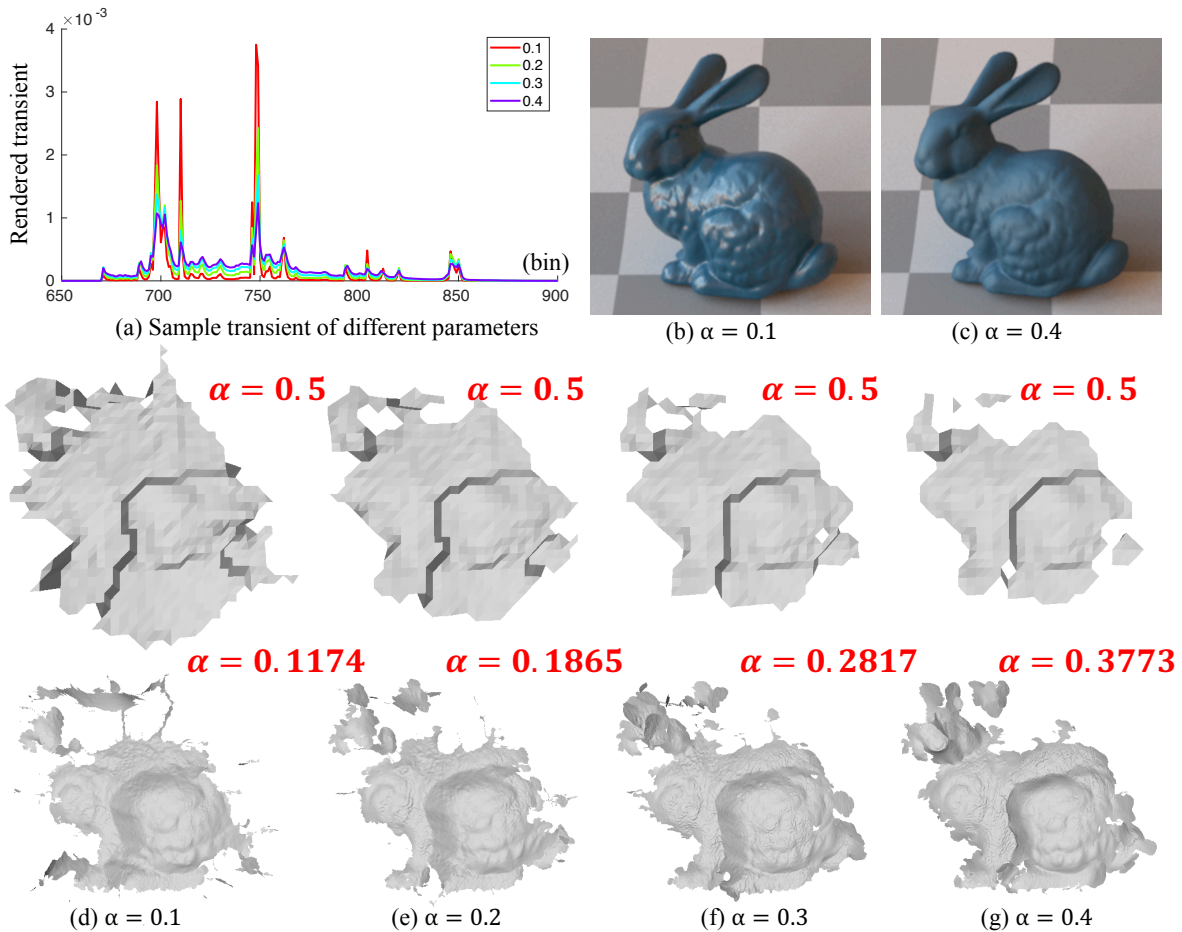


Figure 5.10: **Recovering both shape and reflectance.** (a) We compare transients rendered for different α values to visualize the effect of reflectance on NLOS measurements. (b, c) We also visualize the reflectance by using Mitsuba [56] to render the scene under ambient light. (d - g) We show optimization results for different α values, with the initial shape and α at the top, and the optimized results at the bottom.

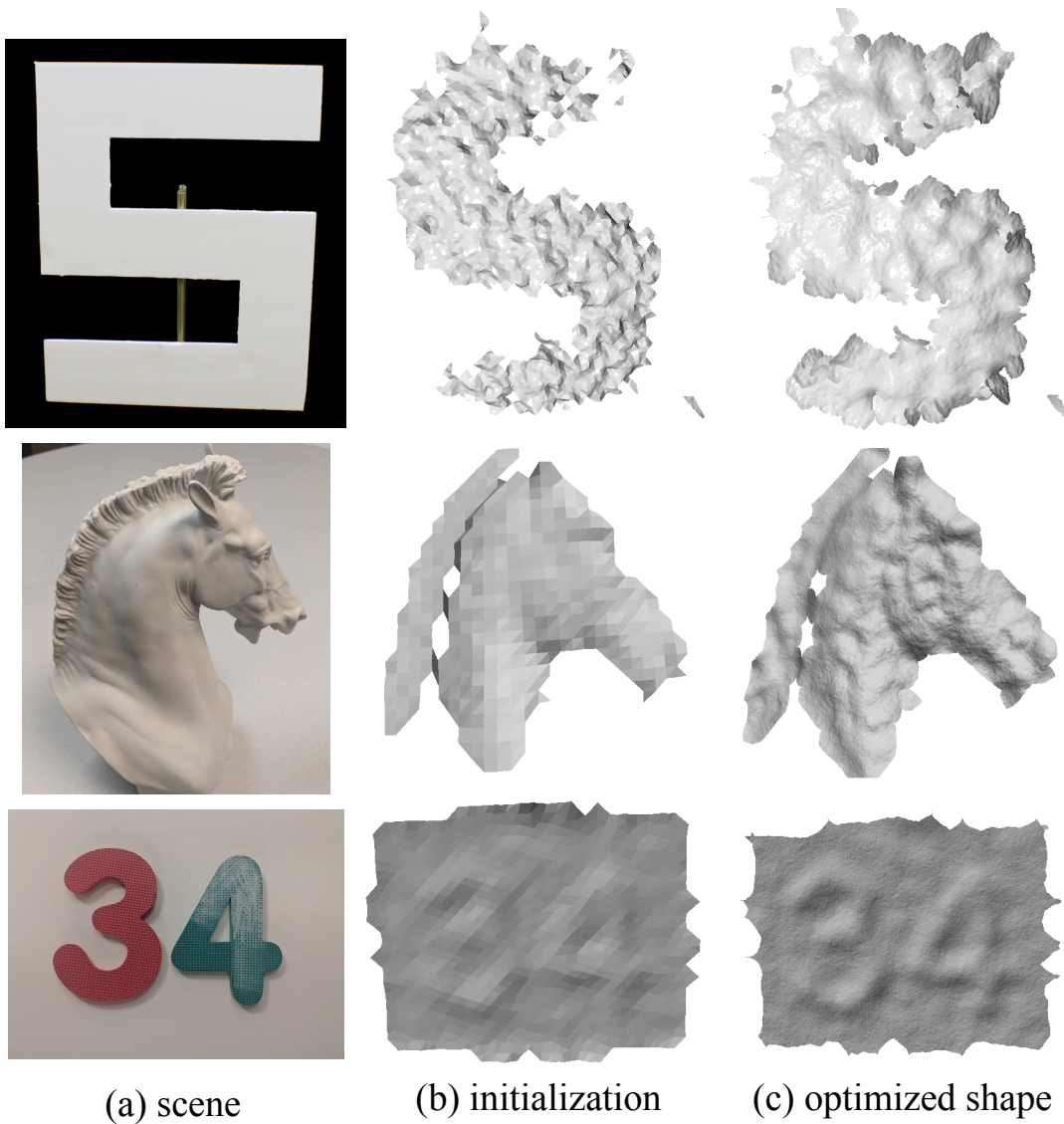


Figure 5.11: **NLOS surface reconstruction using SPAD measurements.** (Top) A diffuse object from [96]. (Middle) A diffuse horse statue. (Bottom) Digit relief on a planar object. In our experiment, we cover the digits with white paper, to increase SNR.

Algorithm 8 Adam routines for optimizing NLOS mesh vertices.

Require: Initial NLOS vertices v_0 .

Require: Initial learning rates η_0 .

Require: Reflectance parameters π .

Require: NLOS triangles T .

Require: Measured transient \tilde{I} .

Require: Virtual source l and detector s .

Require: Regularization weight λ .

Require: Number of iterations N .

Require: Number of rendering samples J .

```
1: function OPTSURFACE( $v_0, \eta_0, \pi, T, \tilde{I}, l, s, \lambda, N, J$ )
2:    $v \leftarrow v_0$ . ▷ Initialize estimate.
3:    $\eta \leftarrow \eta_0$ . ▷ Initialize learning rates.
4:   for  $n \in \{1, \dots, N\}$  do
5:      $\mathcal{S}_{\text{NLOS}} \leftarrow \text{CreateMesh}(v, T)$ .
6:     ▷ Render required quantities.
7:      $I \leftarrow \text{Render}(g; \mathcal{S}_{\text{NLOS}}, \pi, l, s, J)$ .
8:      $I_v \leftarrow \text{Render}(g_s; \mathcal{S}_{\text{NLOS}}, \pi, l, s, J)$ .
9:      $R_v \leftarrow \text{ComputeReguGradient}(\mathcal{S}_{\text{NLOS}})$ .
10:    ▷ Compute gradient.
11:     $g \leftarrow (I - \tilde{I}) I_v + \lambda R_v$ 
12:    ▷ Perform Adam updates.
13:     $v \leftarrow \text{AdamUpdateParameters}(v, \eta, g)$ 
14:     $\eta \leftarrow \text{AdamUpdateLearningRates}(v, \eta, g)$ 
15:  end for
16:  return  $v$ .
17: end function
```

Algorithm 9 Adam routines for optimizing NLOS mesh reflectance.

Require: Initial reflectance parameters π_0 .

Require: Initial learning rates η_0 .

Require: NLOS mesh $\mathcal{S}_{\text{NLOS}}$.

Require: Measured transient \tilde{I} .

Require: Virtual source l and detector s .

Require: Number of iterations N .

Require: Number of rendering samples J .

```

1: function OPTREFLECTANCE( $\pi_0, \eta_0, \mathcal{S}_{\text{NLOS}}, \tilde{I}, l, s, N, J$ )
2:    $\pi \leftarrow \pi_0$ . ▷ Initialize estimate.
3:    $\eta \leftarrow \eta_0$ . ▷ Initialize learning rates.
4:   for  $n \in \{1, \dots, N\}$  do
5:     ▷ Render required quantities.
6:      $I \leftarrow \text{Render}(g; \mathcal{S}_{\text{NLOS}}, \pi, l, s, J)$ .
7:      $I_\pi \leftarrow \text{Render}(g_r; \mathcal{S}_{\text{NLOS}}, \pi, l, s, J)$ .
8:     ▷ Compute gradient.
9:      $g \leftarrow (I - \tilde{I}) I_\pi$ 
10:    ▷ Perform Adam updates.
11:     $\pi \leftarrow \text{AdamUpdateParameters}(\pi, \eta, g)$ 
12:     $\eta \leftarrow \text{AdamUpdateLearningRates}(\pi, \eta, g)$ 
13:  end for
14:  return  $\pi$ .
15: end function

```

Chapter 6

Conclusions and Future Work

We seek to enable 3D estimation techniques for a large class of real-world objects. By leveraging the fact that light propagation in a scene can be interpreted in terms of light paths, which is the path traced by a photon from a light source to its eventual observation at a sensor pixel, we propose a shape estimation framework that utilizes multi-bounce light paths and deals with objects with complex reflectance. We measure how a light ray changes its properties after interacting with the scene and associate the changes to geometric and reflectance information of the scene. Our framework provides great advantages since light propagation in a scene will follow physics law which leads to simpler techniques for understanding scene properties. In particular, our contributions are summarized in Table 6.1.

Bounce	Material	Properties	Physical laws	Application
2	glass	position, direction	Snell's law	Shape estimation for transparent objects [129]
2	opaque	position, direction, ToF	cosine law	Shape estimation for concave objects from two-bounce light paths [130]
3	opaque	position, ToF	Fermat's principle	Non-line-of-sight imaging using first-returning photons [126, 127]
3	opaque	position, ToF, radiance	light transport	Non-line-of-sight imaging using radiance for surface optimization [128]

Table 6.1: **Contributions.** We show in this thesis how using simple physics laws associated with light-object interactions are useful in dealing with multi-bounce light paths and complex materials. We show shape estimation techniques of different scenarios where classical computer vision algorithms cannot handle.

Our research builds toward foundational theories of shape estimation of visually-complex scenes by using light paths as the primitive. By explicitly using simple physical laws to constrain light paths, we open new opportunities in shape estimation for complex scenes.

6.1 Future work

$\langle N, K, M, T \rangle$ analysis. In [74], Kutulakos and Steger show theoretical analysis on determining the tractability of different imaging setup to capture the shape of specular and transparent objects. The characterization involves the number of camera views N , number of bounces in the light path K , and number of reference points M . In this thesis, we seek to expand this analysis to different materials and with ToF measurements T , where $T = 1$ indicates the availability of ToF information and $T = 0$ without.

Chapter 2 is reparameterization of a theory shown in [74], which corresponds to $\langle 3, 2, 2, 0 \rangle$. We show that if the glass object of interest can be approximated with local planar patch, then we can recover the object with a single view, $\langle 1, 2, 2, 0 \rangle$.

In Chapter 3, we establish the sufficient condition for unique shape recovery using two-bounce light paths, this shows $\langle 1, 2, 2, 1 \rangle$ is tractable when the measurements satisfy the sufficient conditions.

In Chapter 4, we use local planarity assumption to link ToF of first-returning photons to estimate points and surface normal belonging to the hidden object, this is attempting to show $\langle 1, 3, 2, 1 \rangle$. However, as shown in Chapter 4, we can only reconstruct sparse points on the surface. Also, the locations of these reconstructed points are determined by the geometry and imaging setup. One possible future direction is to characterize the spatial extent of the reconstruction of the proposed algorithm for simple shapes, such as convex objects.

In Chapter 5, we show another technique of using a surface optimization framework to recover the hidden object, which corresponds to the $\langle 1, 3, 2, 1 \rangle$ setup. However, since our proposed framework relies on a non-convex energy function, it is hard to establish a theoretical guarantee. Based on empirical results, we notice when objects are more spiky, for example, the armadillo scene in Figure 5.5, our optimization framework seems to not perform as good as other types of surface geometry. Our method based on mesh gradient flow may contain fundamental drawbacks in such scenarios.

Light transport aware sensing. In each chapter, we explain the hardware setup for the specific application. However, our ability to efficiently capture information is still limited. In most cases, we need to sweep through every possible pair of light rays and gather the associated properties of light rays, which is time-consuming.

Scenes inherently exhibit some simple characteristics. Similar to the concept mentioned Chapter 4, we use first-returning photons to recover hidden objects. A subset of information contains a lot of information regarding the geometry of the scene, therefore, we don't necessarily need the information of all entries of the 5D light transient to recover the scene. As shown in Chapter 4, we can achieve high SNR for sensing the first-returning photon with a low number of photons. These type of analysis will be beneficial for the design for a more realistic imaging setup that works under different imaging budget.

Surface optimization with geometric constraints. One of the major advantages of studying light paths is that we can find geometric constraints regarding the scene as shown in Chapter 2 - 4. By studying the physical constraints, we are able to locate 3D points on a transparent object (Chap. 2), 3D points on a concave shape object (Chap. 3), and 3D points belonging to the non-

line-of-sight object (Chap. 4). However, most objects in the real world are not merely 3D points, they are surfaces. This is exactly the problem that we attempt to solve in Chapter 5 to directly find a surface that matches the measurements through surface optimizing. The measurements used are the radiometric information from the 5D light transient. A good next step is to use a similar idea of surface optimization on a broader class of measurements of light paths, for example, geometric constraints.

Ideas presented in [20, 57] may lead a good direction in coming up the derivation for surface gradients based on geometric clues. The light path of interest described in Chapter 2 - Chapter 4 are all shortest light paths. More specifically, in Chapter 2 and 4, the light paths are specular paths. Therefore, we may be able to extend ideas of exploring specular paths to exploring a surface that creates the specular paths that we measured.

Blind shape estimation from multi-bounce light paths. In this thesis, we show three scenarios to recover shape from multi-bounce light paths. However, we heavily rely on the prior knowledge of the scene; for instance, we know we are measuring a transparent object and not a concave object. The method proposed in [129] will only work for transparent objects and need to treat other physical phenomena involved, such as interreflection and total reflection, as noise.

A holistic framework to directly infer scene properties from multi-bounce light paths will be more applicable in the real world. All scene properties, including geometry and reflectance, are encoded in attributes of light paths. A broader future work for this thesis is to solve the more general shape and reflectance estimation without prior knowledge of the scene.

Bibliography

- [1] S. Achar, J. R. Bartels, W. L. Whittaker, K. N. Kutulakos, and S. G. Narasimhan. Epipolar time-of-flight imaging. *TOG*, 36(4):37, 2017. 4
- [2] Y. Adato, Y. Vasilyev, O. Ben-Shahar, and T. Zickler. Toward a theory of shape from specular flow. In *ICCV*, 2007. 4
- [3] N. Alldrin, T. Zickler, and D. Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance. In *CVPR*, 2008. 4
- [4] B. Appleton and H. Talbot. Globally minimal surfaces by continuous maximal flows. *IEEE transactions on pattern analysis and machine intelligence*, 28(1):106–118, 2006. 73, 77
- [5] V. Arellano, D. Gutierrez, and A. Jarabo. Fast back-projection for non-line of sight reconstruction. *Optics Express*, 25(10):11574–11583, 2017. 36, 70
- [6] J. Arvo. *Analytic methods for simulated light transport*. PhD thesis, PhD thesis, Yale University, 1995. 76
- [7] P. N. Belhumeur, D. J. Kriegman, and A. L. Yuille. The bas-relief ambiguity. *International Journal of Computer Vision*, 35(1):33–44, 1999. 15
- [8] J. Bertolotti, E. G. van Putten, C. Blum, A. Legendijk, W. L. Vos, and A. P. Mosk. Non-invasive imaging through opaque scattering layers. *Nature*, 491(7423):232, 2012. 36
- [9] J. Boger-Lombard and O. Katz. Non line-of-sight localization by passive optical time-of-flight. *ArXiv e-prints*, Aug. 2018. 36, 37
- [10] T. Bonfort, P. Sturm, and P. Gargallo. General specular surface triangulation. In *ACCV*, 2006. 4
- [11] M. Botsch and L. Kobbelt. A remeshing approach to multiresolution modeling. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pages 185–192, 2004. 84
- [12] K. L. Bouman, V. Ye, A. B. Yedidia, F. Durand, G. W. Wornell, A. Torralba, and W. T. Freeman. Turning corners into cameras: Principles and methods. In *ICCV*, 2017. 36, 37
- [13] T. Brochu and R. Bridson. El topo, 2009. <https://www.cs.ubc.ca/labs/imager/tr/2009/eltopo/eltopo.html>. 84, 85
- [14] T. Brochu and R. Bridson. Robust topological operations for dynamic explicit surfaces. *SIAM Journal on Scientific Computing*, 31(4):2472–2493, 2009. 84

- [15] M. Buttafava, J. Zeman, A. Tosi, K. Eliceiri, and A. Velten. Non-line-of-sight imaging using a time-gated single photon avalanche diode. *Optics express*, 23(16):20997–21011, 2015. [6](#), [35](#), [36](#), [38](#), [58](#), [60](#), [61](#), [66](#), [68](#), [71](#)
- [16] M. Chandraker. What camera motion reveals about shape with unknown brdf. In *CVPR*, 2014. [5](#)
- [17] M. Chandraker, J. Bai, and R. Ramamoorthi. A theory of differential photometric stereo for unknown isotropic brdfs. In *CVPR*, 2011. [5](#)
- [18] M. Chandraker, D. Reddy, Y. Wang, and R. Ramamoorthi. What object motion reveals about shape with unknown brdf and lighting. In *CVPR*, 2013. [5](#)
- [19] E. Charbon. Will avalanche photodiode arrays ever reach 1 megapixel. In *International Image Sensor Workshop*, 2007. [38](#)
- [20] M. Chen and J. Arvo. Theory and application of specular path perturbation. *TOG*, 19(4):246–278, 2000. [102](#)
- [21] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 303–312, 1996. [75](#)
- [22] A. Delaunoy and M. Pollefeys. Photometric bundle adjustment for dense multi-view 3d modeling. In *CVPR*, 2014. [73](#)
- [23] A. Delaunoy and E. Prados. Gradient flows for optimizing triangular mesh-based surfaces: Applications to 3d reconstruction problems dealing with visibility. *IJCV*, 95(2):100–123, 2011. [73](#), [77](#), [78](#), [79](#), [80](#)
- [24] M. Desbrun, M. Meyer, P. Schröder, and A. H. Barr. Implicit fairing of irregular meshes using diffusion and curvature flow. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 317–324, 1999. [73](#)
- [25] I. Dokmanić, Y. M. Lu, and M. Vetterli. Can one hear the shape of a room: The 2-d polygonal case. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 321–324. IEEE, 2011. [8](#)
- [26] J. Dupuy. *Photorealistic Surface Rendering with Microfacet Theory*. PhD thesis, Université Claude Bernard - Lyon I, 2015. [27](#)
- [27] P. Dutré, K. Bala, and P. Bekaert. *Advanced global illumination*. AK Peters, Ltd., 2006. [71](#), [81](#)
- [28] G. Dziuk and C. M. Elliott. Finite elements on evolving surfaces. *IMA journal of numerical analysis*, 27(2):262–292, 2007. [77](#)
- [29] I. Eckstein, J.-P. Pons, Y. Tong, C.-C. Kuo, and M. Desbrun. Generalized surface flows for mesh processing. In *Proceedings of the fifth Eurographics symposium on Geometry processing*, pages 183–192, 2007. [73](#), [77](#)
- [30] G. Fishman. *Monte Carlo: concepts, algorithms, and applications*. Springer Science & Business Media, 1996. [81](#), [82](#), [84](#)

- [31] I. Freund. Looking through walls and around corners. *Physica A: Statistical Mechanics and its Applications*, 168(1):49–65, 1990. [35](#)
- [32] P. Gargallo, E. Prados, and P. Sturm. Minimizing the reprojection error in surface reconstruction from images. In *ICCV*, 2007. [73](#)
- [33] G. Gariepy, N. Krstajić, R. Henderson, C. Li, R. R. Thomson, G. S. Buller, B. Heshmat, R. Raskar, J. Leach, and D. Faccio. Single-photon sensitive light-in-flight imaging. *Nature communications*, 6:6021, 2015. [6](#)
- [34] G. Gariepy, F. Tonolini, R. Henderson, J. Leach, and D. Faccio. Detection and tracking of moving objects hidden from view. *Nature Photonics*, 10(1):23–26, 2016. [36](#)
- [35] A. S. Georghiades. Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo. In *ICCV*, 2003. [4](#)
- [36] A. Geva, Y. Y. Schechner, Y. Chernyak, and R. Gupta. X-ray computed tomography through scatter. In *ECCV*, 2018. [74](#)
- [37] I. Gkioulekas, A. Levin, F. Durand, and T. Zickler. Micron-scale light transport decomposition using interferometry. *TOG*, 34(4):37, 2015. [6](#)
- [38] I. Gkioulekas, A. Levin, and T. Zickler. An evaluation of computational imaging techniques for heterogeneous inverse scattering. *ECCV*, 2016. [74](#)
- [39] I. Gkioulekas, S. Zhao, K. Bala, T. Zickler, and A. Levin. Inverse volume rendering with material dictionaries. *TOG*, 32(6):162, 2013. [5](#), [74](#)
- [40] M. Goesele, H. Lensch, J. Lang, C. Fuchs, and H.-P. Seidel. Disco: acquisition of translucent objects. *TOG*, 23(3):835–844, 2004. [5](#)
- [41] D. B. Goldman, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and spatially-varying brdfs from photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6):1060–1071, 2010. [4](#)
- [42] J. Gregson, M. Krimerman, M. B. Hullin, and W. Heidrich. Stochastic tomography and its applications in 3d imaging of mixing fluids. *TOG*, 31(4):52, 2012. [5](#)
- [43] M. Gupta, A. Agrawal, A. Veeraraghavan, and S. G. Narasimhan. A practical approach to 3d scanning in the presence of interreflections, subsurface scattering and defocus. *International Journal of Computer Vision*, 102(1-3):33–55, 2013. [3](#)
- [44] M. Gupta, S. K. Nayar, M. B. Hullin, and J. Martin. Phasor imaging: A generalization of correlation-based time-of-flight imaging. *TOG*, 34(5):156, 2015. [4](#)
- [45] O. Gupta, T. Willwacher, A. Velten, A. Veeraraghavan, and R. Raskar. Reconstruction of hidden 3d shapes using diffuse reflections. *Optics express*, 20(17):19096–19108, 2012. [70](#)
- [46] E. T. Hale, W. Yin, and Y. Zhang. Fixed-point continuation for ℓ_1 -minimization: Methodology and convergence. *SIAM Journal on Optimization*, 19(3):1107–1130, 2008. [83](#)
- [47] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. [1](#)

- [48] F. Heide, M. B. Hullin, J. Gregson, and W. Heidrich. Low-budget transient imaging using photonic mixer devices. *TOG*, 32(4):45, 2013. 6
- [49] F. Heide, M. O’Toole, K. Zhang, D. B. Lindell, S. Diamond, and G. Wetzstein. Robust non-line-of-sight imaging with single photon detectors. *ArXiv e-prints*, Nov. 2017. 36, 70, 72
- [50] F. Heide, L. Xiao, W. Heidrich, and M. Hullin. Diffuse mirrors: 3d reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors. In *CVPR*, 2014. 36, 70
- [51] Q. Hernandez, D. Gutierrez, and A. Jarabo. A Computational Model of a Single-Photon Avalanche Diode Sensor for Transient Imaging. *ArXiv e-prints*, Feb. 2017. 6, 38, 85, 86
- [52] A. Hertzmann and S. M. Seitz. Example-based photometric stereo: shape reconstruction with general, varying brdfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1254–1264, 2005. 4
- [53] B. Horn and M. Brooks. *Shape from shading*. MIT press, 1989. 1, 8
- [54] Z. Hui and A. C. Sankaranarayanan. Shape and spatially-varying reflectance estimation from virtual exemplars. *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, 39(10):2060–2073, 2017. 4
- [55] C. Inoshita, Y. Mukaigawa, Y. Matsushita, and Y. Yagi. Shape from single scattering for translucent objects. In *ECCV*, 2012. 5
- [56] W. Jakob. Mitsuba renderer, 2010. <http://www.mitsuba-renderer.org>. 27, 96
- [57] W. Jakob and S. Marschner. Manifold exploration: a markov chain monte carlo technique for rendering scenes with difficult specular transport. *TOG*, 31(4):58, 2012. 102
- [58] A. Jarabo, J. Marco, A. Muñoz, R. Buisan, W. Jarosz, and D. Gutierrez. A framework for transient rendering. *TOG*, 33(6):177:1–177:10, 2014. 56
- [59] A. Jarabo, B. Masia, J. Marco, and D. Gutierrez. Recent advances in transient imaging: A computer graphics and vision perspective. *Visual Informatics*, 1(1):65–79, 2017. 5, 35
- [60] A. Kadambi, J. Schiel, and R. Raskar. Macroscopic interferometry: Rethinking depth estimation with frequency-domain time-of-flight. In *CVPR*, 2016. 6
- [61] A. Kadambi, H. Zhao, B. Shi, and R. Raskar. Occluded imaging with time-of-flight sensors. *TOG*, 35(2):15, 2016. 36
- [62] J. T. Kajiya. The rendering equation. In *ACM Siggraph Computer Graphics*, volume 20, pages 143–150, 1986. 70, 81
- [63] J. Kaminski, S. Lowitzsch, M. C. Knauer, and G. HÅd’usler. Full-field shape measurement of specular surfaces. In *Intl. Workshop on Automatic Processing of Fringe Patterns*, 2005. 4, 15
- [64] A. S. Kaplanyan, J. Hanika, and C. Dachsbacher. The natural-constraint representation of the path space for efficient light transport simulation. *ACM Transactions on Graphics (TOG)*, 33(4):102, 2014. 82

- [65] O. Katz, P. Heidmann, M. Fink, and S. Gigan. Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations. *Nature photonics*, 8(10):784, 2014. [36](#), [37](#)
- [66] O. Katz, E. Small, and Y. Silberberg. Looking around corners and through thin turbid layers in real time with scattered incoherent light. *Nature photonics*, 6(8):549–553, 2012. [36](#)
- [67] P. Khungurn, D. Schroeder, S. Zhao, K. Bala, and S. Marschner. Matching real fabrics with micro-appearance models. *TOG*, 35(1):1:1–1:26, 2015. [74](#)
- [68] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *ICLR*, 2015. [75](#), [83](#), [85](#)
- [69] A. Kirmani, T. Hutchison, J. Davis, and R. Raskar. Looking around the corner using ultrafast transient imaging. *IJCV*, 95(1):13–28, 2011. [33](#), [35](#)
- [70] J. Klein, M. Laurenzis, D. L. Michels, and M. B. Hullin. A quantitative platform for non-line-of-sight imaging problems. In *Proceedings BMVC*, pages 1–12, 2018. [82](#)
- [71] J. Klein, C. Peters, J. Martín, M. Laurenzis, and M. B. Hullin. Tracking objects outside the line of sight using 2d intensity images. *Scientific Reports*, 6:32491, 2016. [36](#), [37](#)
- [72] J. J. Koenderink. What does the occluding contour tell us about solid shape? *Perception*, 13(3):321–330, 1984. [77](#)
- [73] K. Kolev and D. Cremers. Integration of multiview stereo and silhouettes via convex functionals on convex domains. In *ECCV*, 2008. [73](#)
- [74] K. N. Kutulakos and E. Steger. A theory of refractive and specular 3d shape by light-path triangulation. *International Journal of Computer Vision*, 76(1):13–29, 2008. [2](#), [4](#), [5](#), [7](#), [10](#), [13](#), [14](#), [15](#), [18](#), [101](#)
- [75] M. La Manna, F. Kine, E. Breitbach, J. Jackson, T. Sultan, and A. Velten. Error backprojection algorithms for non-line-of-sight imaging. *IEEE transactions on pattern analysis and machine intelligence*, 2018. [36](#)
- [76] A. Levis, Y. Schechner, A. Aides, and A. Davis. Airborne three-dimensional cloud tomography. *ICCV*, 2015. [5](#), [74](#)
- [77] F. Li, F. Willomitzer, P. Rangarajan, M. Gupta, A. Velten, and O. Cossairt. Sh-tof: Micro resolution time-of-flight imaging with superheterodyne interferometry. In *IEEE International Conference on Computational Photography (ICCP)*, 2018. [6](#)
- [78] D. Liu, X. Chen, and Y. H. Yang. Frequency-based 3d reconstruction of transparent and specular objects. In *CVPR*, 2014. [5](#)
- [79] H.-T. D. Liu, M. Tao, and A. Jacobson. Paparazzi: Surface editing by way of multi-view image processing. *ACM TOG*, 2018. [73](#), [77](#), [84](#)
- [80] M. Liu, R. Hartley, and M. Salzmann. Mirror surface reconstruction from a single image. In *CVPR*, 2013. [4](#)
- [81] S. Liu, T.-T. Ng, and Y. Matsushita. Shape from second-bounce of light transport. In *ECCV*, 2010. [20](#)
- [82] S. Lombardi and K. Nishino. Radiometric scene decomposition: Scene reflectance, illumi-

- nation, and geometry from rgb-d images. In *3D Vision (3DV), 2016 Fourth International Conference on*, pages 305–313, 2016. [74](#)
- [83] M. M. Loper and M. J. Black. Opendr: An approximate differentiable renderer. In *ECCV*, 2014. [74](#)
- [84] T. Maeda, A. Kadambi, Y. Y. Schechner, and R. Raskar. Dynamic heterodyne interferometry. In *IEEE International Conference on Computational Photography (ICCP)*, 2018. [6](#)
- [85] S. R. Marschner. Inverse rendering in computer graphics, phd thesis. *Ithaca, NY, USA*, 1998. [73](#), [74](#)
- [86] W. Matusik, H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. *TOG*, 22(3):759–769, 2003. [27](#)
- [87] M. Meyer, M. Desbrun, P. Schröder, and A. H. Barr. Discrete differential-geometry operators for triangulated 2-manifolds. In *Visualization and mathematics III*, pages 35–57. 2003. [73](#)
- [88] D. P. Mitchell. Consequences of stratified sampling in graphics. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 277–280. ACM, 1996. [82](#)
- [89] S. G. Narasimhan, M. Gupta, C. Donner, R. Ramamoorthi, S. K. Nayar, and H. W. Jensen. Acquiring scattering properties of participating media by dilution. *TOG*, 25(3):1003–1012, 2006. [5](#)
- [90] S. K. Nayar, K. Ikeuchi, and T. Kanade. Shape from interreflections. *International Journal of Computer Vision*, 6(3):173–195, 1991. [8](#), [20](#)
- [91] 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. *TOG*, 25(3):935–944, 2006. [3](#)
- [92] S. Osher and R. Fedkiw. *Level set methods and dynamic implicit surfaces*, volume 153. Springer Science & Business Media, 2006. [73](#), [75](#)
- [93] M. O’Toole, S. Achar, S. G. Narasimhan, and K. N. Kutulakos. Homogeneous codes for energy-efficient illumination and imaging. *TOG*, 34(4):35, 2015. [3](#), [4](#)
- [94] M. O’Toole, F. Heide, K. Lindell, David B. and Zang, S. Diamond, and G. Wetzstein. Reconstructing transient images from single-photon sensors. In *CVPR*, 2017. [6](#)
- [95] 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. *TOG*, 33(4):87, 2014. [33](#)
- [96] M. O’Toole, D. B. Lindell, and G. Wetzstein. Confocal non-line-of-sight imaging based on the light-cone transform. *Nature*, 555(7696):338, 2018. [36](#), [70](#), [85](#), [86](#), [91](#), [97](#)
- [97] D. Panozzo and A. Jacobson. *LIBIGL: A C++ library for geometry processing without a mesh data structure*. SGP 2014 Graduate School, 2014. [85](#)
- [98] S. G. Parker, J. Bigler, A. Dietrich, H. Friedrich, J. Hoberock, D. Luebke, D. McAllister,

- M. McGuire, K. Morley, A. Robison, et al. Optix: a general purpose ray tracing engine. *ACM TOG*, 29(4):66, 2010. 82, 84
- [99] A. Paszke, S. Chintala, R. Collobert, K. Kavukcuoglu, C. Farabet, S. Bengio, I. Melvin, J. Weston, and J. Mariethoz. Pytorch: Tensors and dynamic neural networks in python with strong gpu acceleration, may 2017. 85
- [100] G. Patow and X. Pueyo. A survey of inverse rendering problems. In *Computer graphics forum*, volume 22, pages 663–687. Wiley Online Library, 2003. 74
- [101] G. Patow and X. Pueyo. A survey of inverse surface design from light transport behavior specification. In *Computer Graphics Forum*, volume 24, pages 773–789. Wiley Online Library, 2005. 74
- [102] A. K. Pediredla, M. Buttafava, A. Tosi, O. Cossairt, and A. Veeraraghavan. Reconstructing rooms using photon echoes: A plane based model and reconstruction algorithm for looking around the corner. In *IEEE International Conference on Computational Photography (ICCP)*, 2017. 36, 58, 67, 71
- [103] A. K. Pediredla and I. Gkioulekas. Mitsuba time-of-flight renderer. <https://github.com/cmu-ci-lab/MitsubaToFRenderer>, 2018. 27
- [104] C. Peters, J. Klein, M. B. Hullin, and R. Klein. Solving trigonometric moment problems for fast transient imaging. *TOG*, 34(6):220, 2015. 6
- [105] G. T. Pfeiffer and Y. Sato. On stochastic optimization methods for monte carlo least-squares problems. *arXiv preprint arXiv:1804.10079*, 2018. 83
- [106] M. Pharr, W. Jakob, and G. Humphreys. *Physically based rendering: From theory to implementation*. Morgan Kaufmann, 2016. 71, 81
- [107] M. Pharr, C. Kolb, R. Gershbein, and P. Hanrahan. Rendering complex scenes with memory-coherent ray tracing. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 101–108, 1997. 82
- [108] A. Pilleboue, G. Singh, D. Coeurjolly, M. Kazhdan, and V. Ostromoukhov. Variance analysis for monte carlo integration. *ACM Transactions on Graphics (TOG)*, 34(4):124, 2015. 82
- [109] J.-P. Pons, R. Keriven, and O. Faugeras. Modelling dynamic scenes by registering multi-view image sequences. In *CVPR*, 2005. 73
- [110] R. Ramesh and J. Davis. 5d time-light transport matrix: What can we reason about scene properties? *Tech. rep., MIT*, 2008. 33
- [111] A. C. Sankaranarayanan, A. Veeraraghavan, O. Tuzel, and A. Agrawal. Image invariants for smooth reflective surfaces. In *ECCV*, 2010. 4
- [112] A. C. Sankaranarayanan, A. Veeraraghavan, O. Tuzel, and A. Agrawal. Specular surface reconstruction from sparse reflection correspondences. In *CVPR*, 2010. 4
- [113] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *CVPR*, 2003. 1
- [114] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evalu-

- ation of multi-view stereo reconstruction algorithms. In *CVPR*, 2006. [73](#), [77](#)
- [115] S. M. Seitz, Y. Matsushita, and K. N. Kutulakos. A theory of inverse light transport. In *ICCV*, 2005. [3](#)
- [116] G. S. Settles. *Schlieren and shadowgraph techniques*. Springer, 2001. [5](#)
- [117] S. Shrestha, F. Heide, W. Heidrich, and G. Wetzstein. Computational imaging with multi-camera time-of-flight systems. *ACM Trans. Graphics*, 35(4):33, 2016. [6](#)
- [118] B. M. Smith, M. O’Toole, and M. Gupta. Tracking multiple objects outside the line of sight using speckle imaging. In *CVPR*, 2018. [36](#), [37](#)
- [119] J. E. Solem and N. Chr. A geometric formulation of gradient descent for variational problems with moving surfaces. In *Scale-Space*, pages 419–430. Springer, 2005. [73](#)
- [120] O. Sorkine, D. Cohen-Or, Y. Lipman, M. Alexa, C. Rössl, and H.-P. Seidel. Laplacian surface editing. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pages 175–184, 2004. [73](#)
- [121] Q. Sun, X. Dun, Y. Peng, and W. Heidrich. Depth and transient imaging with compressive spad array cameras. In *CVPR*, 2018. [6](#)
- [122] M. Tarini, H. P. Lensch, M. Goesele, and H.-P. Seidel. 3d acquisition of mirroring objects using striped patterns. *Graphical Models*, 67(4):233–259, 2005. [4](#)
- [123] The CGAL Project. *CGAL User and Reference Manual*. CGAL Editorial Board, 4.13 edition, 2018. [85](#)
- [124] C. Thrampoulidis, G. Shulkind, F. Xu, W. T. Freeman, J. H. Shapiro, A. Torralba, F. N. C. Wong, and G. W. Wornell. Exploiting occlusion in non-line-of-sight active imaging. *IEEE Transactions on Computational Imaging*, 4(3):419–431, 2018. [36](#), [37](#)
- [125] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, 9(2):137–154, 1992. [1](#)
- [126] C.-Y. Tsai, K. N. Kutulakos, S. G. Narasimhan, and A. C. Sankaranarayanan. The geometry of first-returning photons for non-line-of-sight imaging. In *CVPR*, 2017. [35](#), [36](#), [100](#)
- [127] C.-Y. Tsai, A. K. Pediredla, A. Veeraraghavan, K. N. Kutulakos, S. Narasimhan, and A. C. Sankaranarayanan. First-returning photons for non-line-of-sight imaging. *under submission*, 2019. [35](#), [36](#), [100](#)
- [128] C.-Y. Tsai, A. C. Sankaranarayanan, and I. Gkioulekas. Beyond volumetric albedo — a surface optimization framework for non-line-of-sight imaging. *under submission*, 2019. [36](#), [72](#), [100](#)
- [129] C.-Y. Tsai, A. Veeraraghavan, and A. C. Sankaranarayanan. What does a light ray reveal about a transparent object ? In *IEEE International Conference on Image Processing (ICIP)*, 2015. [10](#), [100](#), [102](#)
- [130] C.-Y. Tsai, A. Veeraraghavan, and A. C. Sankaranarayanan. Shape and reflectance from two-bounce light transients. In *IEEE International Conference on Computational Photography (ICCP)*, 2016. [19](#), [20](#), [100](#)

- [131] E. Veach. *Robust monte carlo methods for light transport simulation*. Stanford University Stanford, 1998. [71](#), [81](#)
- [132] E. Veach and L. Guibas. Bidirectional estimators for light transport. In *Photorealistic Rendering Techniques*, pages 145–167. Springer, 1995. [81](#)
- [133] E. Veach and L. J. Guibas. Optimally combining sampling techniques for monte carlo rendering. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 419–428. ACM, 1995. [81](#)
- [134] E. Veach and L. J. Guibas. Metropolis light transport. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 65–76. ACM Press/Addison-Wesley Publishing Co., 1997. [81](#)
- [135] 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*, 3:745, 2012. [33](#), [35](#), [36](#), [70](#), [72](#)
- [136] A. Velten, D. Wu, A. Jarabo, B. Masia, C. Barsi, C. Joshi, E. Lawson, M. Bawendi, D. Gutierrez, and R. Raskar. Femto-photography: capturing and visualizing the propagation of light. *TOG*, 32(4):44, 2013. [6](#)
- [137] I. Wald, P. Slusallek, C. Benthin, and M. Wagner. Interactive rendering with coherent ray tracing. In *Computer graphics forum*, volume 20, pages 153–165. Wiley Online Library, 2001. [82](#)
- [138] I. Wald, S. Woop, C. Benthin, G. S. Johnson, and M. Ernst. Embree: a kernel framework for efficient cpu ray tracing. *ACM TOG*, 33(4):143, 2014. [82](#), [84](#)
- [139] B. Walter, S. R. Marschner, H. Li, and K. E. Torrance. Microfacet models for refraction through rough surfaces. In *Proceedings of the 18th Eurographics conference on Rendering Techniques*, pages 195–206. Eurographics Association, 2007. [75](#)
- [140] T. Wang, M. Chandraker, A. A. Efros, and R. Ramamoorthi. Svbrdf-invariant shape and reflectance estimation from a light-field camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(3):740–754, 2018. [5](#)
- [141] M. Weinmann, A. Osep, R. Ruiters, and R. Klein. Multi-view normal field integration for 3d reconstruction of mirroring objects. In *ICCV*, 2013. [4](#)
- [142] G. Wetzstein, D. Roodnick, W. Heidrich, and R. Raskar. Refractive shape from light field distortion. In *ICCV*, 2011. [5](#), [10](#)
- [143] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):139–144, 1980. [1](#), [8](#)
- [144] F. Xu, G. Shulkind, C. Thrampoulidis, J. H. Shapiro, A. Torralba, F. N. Wong, and G. W. Wornell. Revealing hidden scenes by photon-efficient occlusion-based opportunistic active imaging. *Optics express*, 26(8):9945–9962, 2018. [36](#), [37](#)
- [145] T. Xue, M. Rubinstein, N. Wadhwa, A. Levin, F. Durand, and W. T. Freeman. Refraction wiggles for measuring fluid depth and velocity from video. In *ECCV*, 2014. [5](#)

- [146] A. Yezzi and S. Soatto. Stereoscopic segmentation. *IJCV*, 53(1):31–43, 2003. [73](#)
- [147] K.-J. Yoon, E. Prados, and P. Sturm. Joint estimation of shape and reflectance using multiple images with known illumination conditions. *IJCV*, 86(2):192–210, 2010. [73](#)
- [148] Z. Zhang. Microsoft kinect sensor and its effect. *IEEE multimedia*, 19(2):4–10, 2012. [1](#)
- [149] S. Zhao, L. Wu, F. Durand, and R. Ramamoorthi. Downsampling scattering parameters for rendering anisotropic media. *ACM TOG*, 35(6):166:1–166:11, 2016. [74](#)