3D Reconstruction Using Differential Imaging

Shumian Xin

CMU-RI-TR-22-73

December 2022

Robotics Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Ioannis Gkioulekas, Co-chair
Srinivasa Narasimhan, Co-chair
Keenan Crane
Gordon Wetzstein, Stanford University

Keywords: 3D Reconstruction, Differential Imaging

Abstract

3D reconstruction is at the heart of many computer vision applications, including autonomous driving, visual inspection in manufacturing, and augmented and virtual reality (AR/VR). Because monocular 3D sensing is fundamentally ill-posed, many reconstruction techniques achieve high accuracy by using multiple captures to solve the inverse problem. Depending on the amount of change in these captures relative to the scale of the scene, we can broadly categorize imaging techniques into two groups: non-differential imaging and differential imaging. For example, a stereo system with a large baseline is non-differential, whereas one with a tiny baseline is differential.

Differential imaging offers a few advantages. On the hardware side, because of the tiny changes in measurements, differential imaging systems can be made compact and portable. There are commercially-available sensors at our disposal that already facilitate differential imaging, such as light field cameras and dual-pixel cameras that capture images of a scene under differentially-varying viewpoints. On the algorithm side, differential imaging makes it possible to locally linearize originally nonlinear phenomena so that inverse problems become easier to solve.

In this thesis, we leverage differential imaging to solve three challenging reconstruction problems. First, we introduce a method for non-line-of-sight (NLOS) imaging, an imaging scenario where the scene of interest is not directly visible to the camera. We apply differential imaging by densely scanning a visible surface using a transient imaging system. We then extract a geometric feature that we call the Fermat paths (defined as light paths that satisfy Fermat's principle) from each transient measurement of photons bouncing between the visible surface and the NLOS object. Using the collection of Fermat paths at all scan points, we apply tools from differential geometry to conduct differential analysis and reconstruct the surface of the NLOS objects.

Second, we introduce a method for reconstructing purely specular objects. Our setup illuminates the unknown object with a near-field point light source, and images it with a differentially translating camera. The interaction of light with specular surfaces, specular reflection, is a consequence of Fermat's principle. Therefore, our method for specular shape reconstruction also leverages the theory of Fermat paths. We examine both the geometric and radiometric information in these paths to show that it is possible to uniquely reconstruct the unknown specular object.

Third and last, we introduce a method for single-shot depth from defocus. This is a

fundamentally ill-posed problem when using a conventional camera. To address this, we propose to use a commercially-available dual-pixel sensor, which emulates a stereo system with a differential baseline. We study the image formation model of a dual-pixel camera in the presence of defocus blur, and propose a method to simultaneously estimate the defocus map and the latent all-in-focus image from a single dual-pixel capture.

We hope this thesis will inspire the use of more differential imaging hardware systems and algorithms for 3D reconstruction. The techniques we introduce in this thesis, especially the theory of Fermat paths, will also apply to other domains, including wavefront sensing, acoustic and ultrasonic imaging, lensless imaging, and seismic imaging.

Acknowledgments

I will forever treasure my PhD experience at Carnegie Mellon University. I am deeply grateful to all the amazing people I have met during my time at CMU.

First of all, I would like to express my sincerest gratitude to my PhD advisors: Prof. Ioannis Gkioulekas and Prof. Srinivasa Narasimhan, for their guidance, support, and sincere care of my overall well-being. They are the best PhD advisors and academic role models to work with and to learn from. Srinivas was the instructor of the very first course I took at CMU when I was a master student. He introduced me to the world of computer vision with his intuitive way of teaching, and painted an intriguing picture of all the potential applications that computer vision can enable. I was attracted to doing computer vision research because of him. Thank you Srinivas, for your broad vision so that I can put my PhD research into context, see how our work fills the gap, and always keep the big picture in mind. As one of Yannis' first PhD students, I felt like we started a new journey together back in 2017—me as a PhD student, and him as an assistant professor. Yannis guided me in every aspects of conducting research, from literature research and simulations to optical experiments and paper writing. Every time after I met with Yannis, I left the meeting with a clearer mind of what stage we are in and what directions we should follow to further explore the research problem. Thank you Yannis, for showing me what rigorous academic research is and always steering me to whatever resources I need to make progress.

I am grateful to my committee members: Prof. Keenan Crane and Prof. Gordon Wetzstein. Keenan was the instructor of the Discrete Differential Geometry course—one of my favorite courses at CMU. During the course, I learned a lot about geometry through all the manipulations we did to the Stanford Bunny. Thank you Keenan, for teaching me geometry and for all these beautiful slides you have made and your Youtube videos to make geometry even more attractive. Gordon is a researcher that I look up to. His work is always inspiring. Thank you Gordon, for serving on my committee and providing your insightful opinion.

I am fortunate to have collaborated with Sotiris Nousias, Prof. Kiriakos Kutulakos, and Prof. Aswin Sankaranarayanan on our CVPR 2019 work. Thank you Sotiris, for all the data you helped collect so that we have impressive real-world results to present in the paper. Thank you Kyros, for your detailed guidance on my CVPR oral presentation. I still remember where to emphasize or to pause every time when I give a presentation

on the work. Aswin opened the door of his lab to me when I was a master student at CMU. Thank you Aswin, for being my first research advisor and leading me to the field of computational imaging.

My research internship experience at Google was a wonderful opportunity for me to work with amazing researchers in industry. I would like to thank my hosts and collaborators: Rahul Garg, Neal Wadhwa, Tianfan Xue, Jonathan Barron, Pratul Srinivasan, and Jiawen Chen, for their guidance and insightful discussions which made our ICCV 2021 publication possible. Although the internship was virtual (unfortunately due to COVID), the experience was real.

I would like to thank all my labmates from Srinivas', Yannis', and Aswin's labs for their companion. Thank you Robert, Supreeth, Joe, Tiancheng, Suren, Dinesh, Mark, Adithya, Bowei, Tianyuan, Alankar, Bruce, Bakari, Arjun, Jian, Chia-Yin, Vishwa, Zhuo, Rick, Yi, Anqi, and Byeongjoo, for making my graduate lift at CMU a super fun one.

Last but not least, I would like to thank my parents and my husband Chao Liu for their love and support. Thank you Mom and Dad, for raising me with love and encouragement to shape who I am today. Thank you Chao, for being my best friend, understanding all my emotions, and always giving me support when I need it the most.

To all my loved ones

Contents

List of Figures

List of Tables

Chapter 1

Introduction

3D reconstruction is at the heart of many computer vision applications, including autonomous driving, robot navigation, visual inspection in manufacturing, as well as augmented and virtual reality (AR/VR). Because 3D reconstruction from a single image is fundamentally ill-posed, many techniques aiming for accurate reconstruction use multiple captures to solve the inverse problem. There are a series of ways to take multiple images by changing one or more factors in the imaging process, e.g., a change of illumination in photometric stereo, movement of a dynamic scene in optical flow, a change of viewpoint in stereo and multiview geometry, and a change of optical power in depth from focus/defocus. Depending on the amount of change relative to the scale of the scene, we can categorize 3D reconstruction methods into two groups: non-differential imaging and differential imaging.

Differential imaging makes small changes, which offers benefits and at the same time poses challenges when compared to their non-differential counterpart. The goal of this thesis to leverage differential imaging to develop systems and algorithms for challenging 3D reconstruction problems. In Section 1.1, we introduce already-existing differential imaging systems, both in nature and in computer vision. In Section 1.2, we discuss the advantages and disadvantages of differential imaging. In Section 1.3, we provide a thesis overview and introduce the three reconstruction problems we have explored using differential imaging.

(a) what a fly sees (b) what a jumping spider sees

Figure 1.1: **Differential imaging systems in nature.** Differential imaging mechanisms can be found in many living creatures. (a) Many insects, such as flies, have compound eyes which allow them to simultaneously record their surroundings from thousands of slightly different viewpoints. Thus, they possess a strong capability for detecting fast motions. (b) Jumping spiders have a compact stack of translucent retinas in their eyes. Images formed on each layer exhibit different amounts of defocus blur and are exploited by jumping spiders for depth sensing.

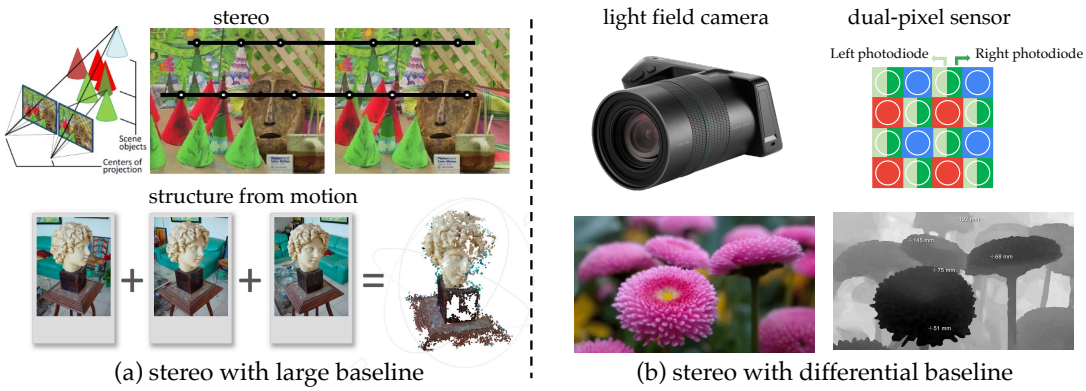
1.1 Differential Imaging Systems

Differential imaging systems can be found both in nature and in computer vision.

Differential imaing systems in nature. Humans perceive the 3D world with a binocular visual system. We can, however, still perceive 3D structure with one eye closed. In this monocular case, a number of visual cues are still in effect. One of them is the scanning mechanism with tiny eye movements [127, 150, 151]. The change in images formed on the retina, termed motion parallax, provides us with a sense of depth—objects that are farther away moves slower than those that are closer [50].

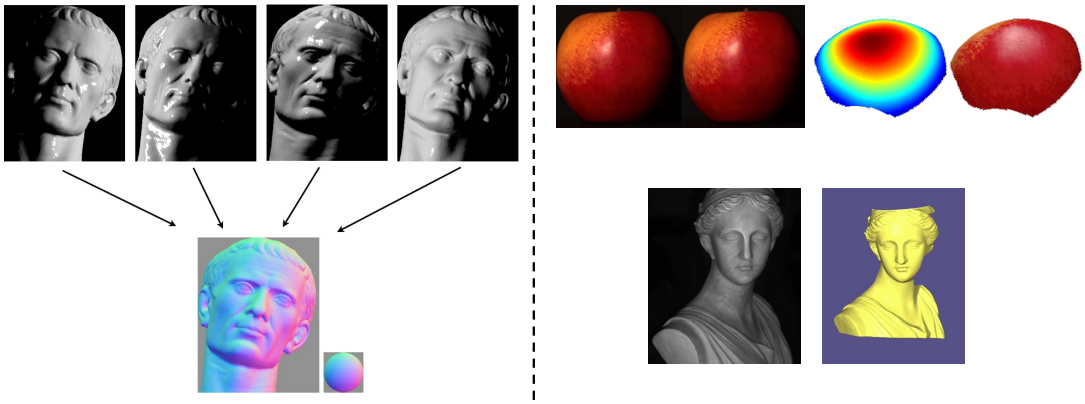
Differential imaging mechanisms can be found in other living creatures as well, as shown in Figure 1.1. Many insects, such as flies, have compound eyes which allow them to simultaneously record their surroundings from thousands of slightly different viewpoints. Thus, they possess a strong capability for detecting fast motions. Jumping spiders have a compact stack of translucent retinas in their eyes. Images formed on each layer exhibit different amounts of defocus blur and are exploited by jumping spiders for depth sensing.

Differential imaging in computer vision. Almost all 3D imaging techniques have differential versions. As shown in Figure 1.2, by changing camera locations, multiview 3D reconstruction algorithms, such as stereo, multiview geometry, and structure from motion, establish correspondences across images to triangulate 3D structure. Correspond-



(a) stereo with large baseline (b) stereo with differential baseline

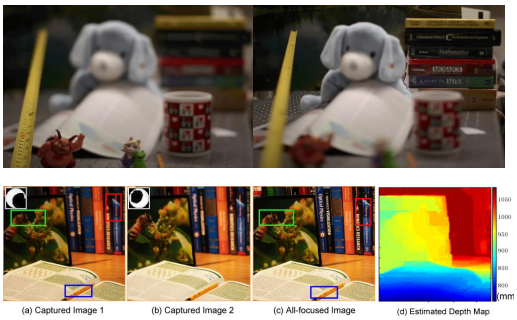
Figure 1.2: **Multiview systems with large and small camera baselines.** By changing camera locations, 3D reconstruction algorithms, such as stereo, multiview geometry, and structure from motion, establish correspondences across images to triangulate 3D structure. Correspondingly, differential imaging in this case refers to imaging systems with a tiny baseline.



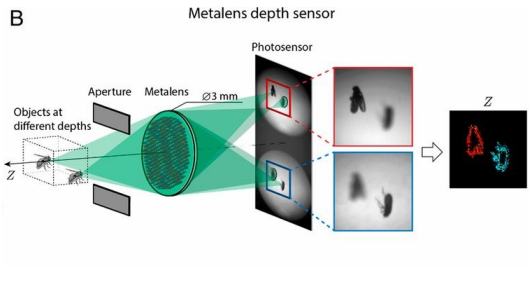
(a) photometric stereo (b) differential photometric stereo

Figure 1.3: **Photometric stereo and differential photometric stereo.** In active imaging systems, such as photometric stereo, one can also control the light source and use the shading information captured under varying illumination to infer surface shape. Differential photometric stereo works in a similar way, but the movement of the light source is very small compared to the scale of the scene. Image courtesy of [33, 110].

ingly, differential imaging in this case refers to imaging systems with tiny baselines, such as multi-camera arrays [186], light field cameras [104], and the recent innovations of dual-pixel (DP) and quad-pixel(QP) sensors in consumer electronics, which can be seen as two-sample and four-sample light field cameras, respectively.



(a) depth from defocus



(b) depth from differential defocus

Figure 1.4: **Depth from defocus and depth from differential defocus.** Depth from defocus methods capture images with distinctive focus settings or camera apertures. On the other hand, depth from differential defocus is inspired by the eys of jumping spiders, and instead explores differential constraints for depth estimation. Image courtesy of [60, 194].

As shown in Figure 1.3, in active imaging systems, such as photometric stereo, one can also control the light source and use the shading information captured under varying illumination to infer surface shape. Differential photometric stereo works in a similar way, but the movement of the light source is tiny compared to the scale of the scene [33, 110].

In addition to changing camera or light source locations, there are also methods that use images captured with different camera intrinsic parameters. For example, as shown in Figure 1.4, depth from defocus methods capture images with distinctive focus settings [48, 141, 173, 185] or camera apertures [194]. On the other hand, inspired by the eyes of jumping spiders, depth from differential defocus techniques work by making a slight change to the aperture size, the optical power, the lens location, or a combination of these factors [6, 7, 60].

1.2 Motivation and Challenges

In this section, we discuss the advantages and disadvantages of differential imaging in comparison to non-differential imaging, which we summarize in Table 1.1.

	differential imaging	non-differential imaging
form factor	compact and portable	relatively large form factor
local linear approximation	possible	impossible
higher-order surface information	computable	unavailable
finding correspondence	easy	hard
scene coverage	small	large
estimation uncertainty	large	small
robustness to noise	less robust	robust

Table 1.1: **Comparison of differential imaging and non-differential imaging.**

1.2.1 Motivation

We are motivated to use differential imaging because of the following benefits it offers.

Compactness and portability. On the hardware side, because differential imaging systems make tiny changes, they can be made more compact and portable. There are commercially available sensors at our disposal that already facilitate differential imaging, including light field cameras, and the recent innovation of dual-pixel (DP) and quad-pixel sensors which can be seen as two-sample and four-sample light field cameras, respectively.

Enbabling local linear approximation. Differential imaging makes it possible to locally linearize originally nonlinear phenomena so that 3D reconstructon problems become easier to solve. Prior works have addressed the problem of recovering various scene properties under differential changes in illumination, camera, and scene, by solving a linearized system. For example, optical flow methods aim to estimate a dense motion field between two image frames taken at small time intervals [16, 77, 115]. Assuming the motion is small, these methods linearize the brightness constancy constraint and solve for a linear system at each pixel. Chandraker and collaborators extensively explore the scene information available with differential motions of illumination, camera, and scene, for various reflectance and camera models with local linear approximation [30, 31, 32, 33, 34, 35, 183].

Obtaining higher-order 3D shape information. As we will see soon in Chapter 2, by adopting differential imaging and performing differential analysis on 3D geometry, one can further obtain higher-order 3D shape information such as normals and curvatures.

Easier correspondence. Differential imaging also makes it easier to establish correspondences across measurements. Because of the tiny change, correspondences across im-

ages can usually be found in a local window rather than a global search within the entire image.

1.2.2 Challenges

Although differential imaging offers the aforementioned benefits, we have to be aware of its challenges.

Small scene coverage. Because of the tiny change made in differential imaging systems, the scene coverage is small compared to its non-differential counterparts.

Larger estimation uncertainty. Differential imaging brings in larger estimation uncertainty. For example, in triangulation-based 3D reconstruction methods, there is an inverse correlation between the baseline and the estimation uncertainty. The same is true for other types of differential imaging systems as well. As the change reduces, redundancy becomes higher across multiple captures. In the extreme case of a zero change, additional measurements impose no further constraints to the 3D scene and the problem boils down to monocular 3D sensing.

Less robustness to noise. Differential analysis usually involves computing spatial or temporal derivatives. This differentiation procedure will amplify high-frequency components as well as noise.

1.3 Thesis Overview

In this thesis, we leverage differential imaging to solve three challenging 3D reconstruction problems, including non-line-of-sight imaging, specular object reconstruction, and single-shot depth from defocus and defocus deblurring.

The wide spectrum of reconstruction problems allow us to explore various aspects of differential imaging. We have applied differential imaging in highly-controlled lab settings with active imaging setups and also in the wild passively using a commercial smartphone camera. The differential imaging scheme adopted in all tasks can be classified as spatial scanning, implemented by differentially translating the sensor, the light source, or both co-locatedly. The number of scan points ranges from millions, thousands, to as few as two.

We have examined geometric and/or radiometric information of light traversed in the scene. For problems exploiting geometric information, differential imaging allows

us to use well-established tools in differential geometry to conduct higher-order analysis on the shape of interest. For problems exploiting radiometric information, we model the image formation and propose ways to invert the process.

We hope this thesis will contribute towards the goal of 3D reconstruction in the wide, especially on systems with limited form-factors, in applications where real-time performance is required, and in scenarios where higher-order surface information is needed. Our specific contributions are listed as follows and are summarized in Figure 1.5.

1.3.1 A Theory of Fermat Paths

Light paths, defined as trajectories of photons traversed in the scene, are determined by the object's geometry and reflectance. For surfaces of aribtrary shapes and general reflectance, the space of paths is infinite. Among all the paths that light can travel, the ones that satisfy Fermat's principle are particularly informative and encode rich geometric information of the surface. We call them the Fermat paths.

In Chapter 2, we present the theory of Fermat paths by making connections to the well-knowned signed distance function (SDF) and show that Fermat paths are a superset of the shortest path. Using tools from differential geometry, we analyze the higher-order properties of the lengths of Fermat paths. Based on this theory, we propose shape reconstruction algorithms from a collection of differentially measured attributes of Fermat paths, including pathlengths or path directions.

1.3.2 Fermat Paths for Non-line-of-sight Imaging

Non-line-of-sight (NLOS) imaging is the problem of reconstructing the shape of one or more objects that are completely occluded and outside the line of sight of both the sensor and the light source. The goal is to perform this reconstruction through indirect observations of the objects, through reflections on intermediate surfaces such as walls or the ground. For example, in the common so-called "looking around the corner" setting (Figure 1.5(a)), an NLOS object is reconstructed through the photons that bounce three times between the source and the sensor, first on the wall, then on the object, and then on the wall once again. This problem was introduced to vision and graphics communities in 2009, by researchers from Ramesh Raskar's group at MIT [91]. They proposed using a time-of-flight sensor, instead of traditional intensity camera, which makes it possible

to separate the three-bounce photons that reflect on the NLOS object, from the single-bounce photons return to the sensor after reflecting on the visible surface. This first paper spurred a new research area, where various groups have been coming up with better time-of-flight imaging systems for capturing these three-bounce photons, as well as better algorithms for reconstructing NLOS scenes from them.

In Chapter 3, we propose a purely geometric solution to NLOS imaging based on theory of Fermat paths. In the context of NLOS imaging, the theory states that a significant amount of photons will travel along Fermat paths between the visible surface and the NLOS object. We apply differential imaging by densely scanning the visible surface using a trasient imaging system comprising a laser and a time-of-flight (ToF) sensor. We propose an approach to extract Fermat paths information from these transient measurements, and apply the shape reconstruction algorithm developed in Chapter 2 to reconstruct high-resolution shapes of objects with arbitrary reflectance in an NLOS setting as if the objects were directly visible to the camera, even though the camera never actually sees them.

1.3.3 Fermat Paths for Specular Object Reconstruction

3D reconstruction of purely specular mirror-like objects has important applications in automated inspection of optical mirrors [109], metal components [154], and astronomical telescopes [158]. However, reconstructing specular objects is often considered challenging in computer vision because they violate the common assumption of Lambertian reflectance in well-established techniques, including photometric stereo [19, 41] and multiview geometry [64]. Therefore, their reconstruction requires algorithms tailored for specular objects, which are known as *shape from specular reflection* [22, 78, 97, 119, 128, 130, 153, 198] or *deflectometry* [28, 94, 154, 190].

Differential imaging is commonly used in reconstucting mirror surfaces, implemented by using light field cameras or wavefront sensors [119, 195]. In our work, we consider an equivalent imaging setup with a differentially-translating camera. The need for dense measurements is because specular surfaces only reflect light in the mirror reflecting direction, in contrast to Lambertian objects which scatter light in all directions. Therefore, capturing these reflections can be challenging in the first place and reflections could be missed if they are not densely measured.

Because the law of specular reflection is derived from Fermat's principle, all light

paths reflecting off a purely specular object are Fermat paths. Once they get measured, we can again apply the theory of Fermat paths for specular shape reconstruction. We show that the reconstruction using only the geometric information of Fermat paths is a one-parameter family of surfaces. We propose to further resolve the remaining ambiguity using the radiometric information of specularities. We detail this work in Chapter 4.

1.3.4 Defocus Map Estimation and Deblurring from a Single Dual-Pixel Image

Depth from defocus techniques have a long history in computer vision. Most of them require multiple captures at different focus settings, which can be impractical for dynamic scenes [48, 141, 173, 185]. Other techniques use specialized optics, e.g., coded apertures, to enable single-shot depth from defocus, overcoming the ambiguities inherent in this problem with a conventional camera [43, 101, 102]. We provide a simple alternative solution by using off-the-shelf dual-pixel(DP) sensors. These sensors were introduced in commercial cameras, such as DSLRs and smartphones to improved autofocus, and nowadays they are becoming increasingly commonplace [2, 74]. Each pixel of a DP sensor is split into two halves, and thus the sensor simultaneously captures two sub images per exposure. As a result, it can be seen as a two-sample light field camera [129] or a stereo system with a differential baseline.

In Chapter 5, We propose a method that takes as input a single dual-pixel image, and simultaneously estimates the image's defocus map—the amount of defocus blur at each pixel—and recovers an all-in-focus image. Prior works have solved the two recovery problems independently of each other, and often require large labeled datasets for supervised training. By contrast, we show that it is beneficial to treat these two closely-connected problems simultaneously. To this end, we set up an optimization problem that, by carefully modeling the optics of dual-pixel images, jointly solves both problems. We use data captured with a consumer smartphone camera to demonstrate that, after a one-time calibration step, our approach improves upon prior works for both defocus map estimation and blur removal, despite being entirely unsupervised.

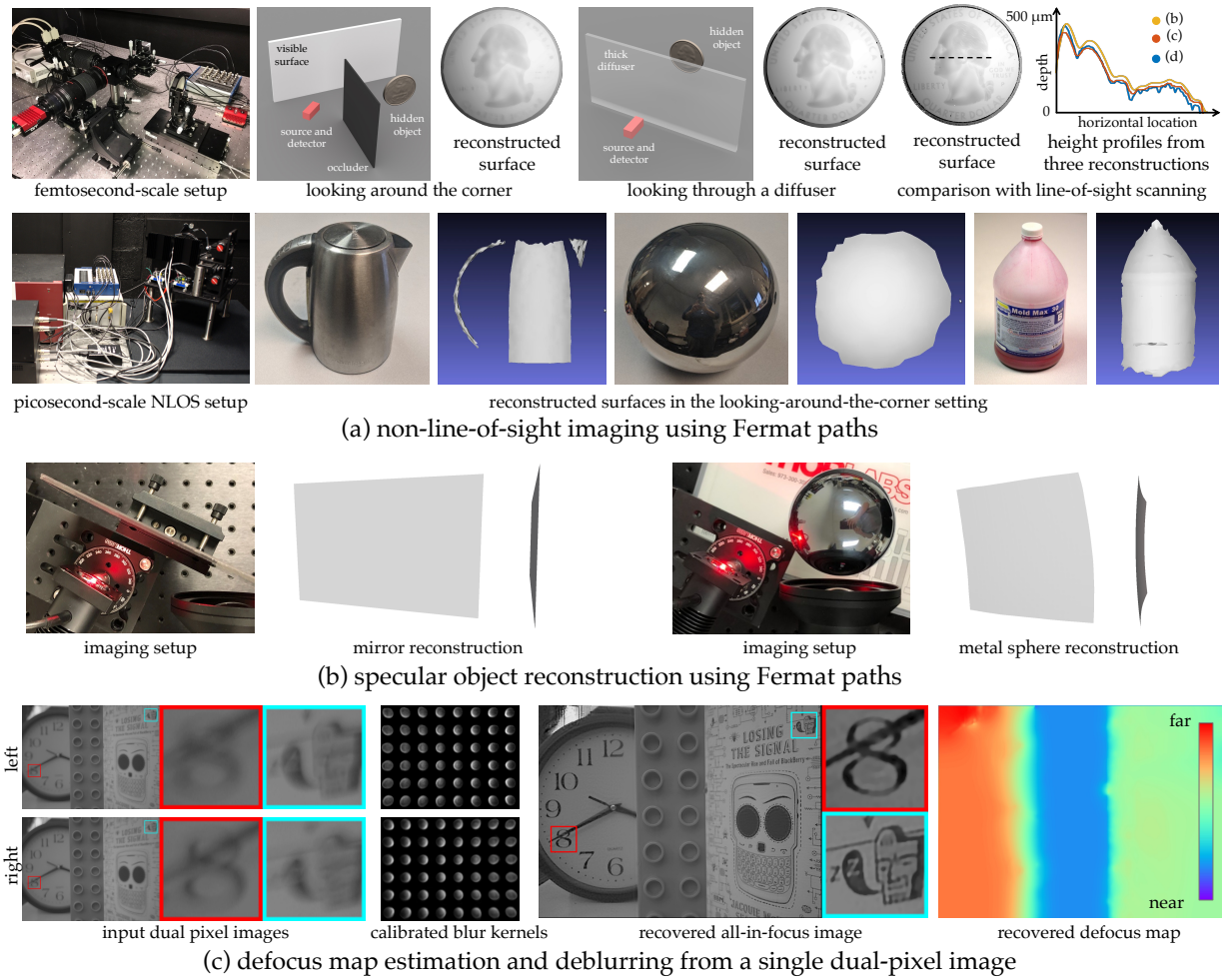


Figure 1.5: **Thesis contributions.** We apply differential imaging to solve three challenging 3D reconstruction problems. (a) Non-line-of-sight (NLOS) imaging aims to reconstruct hidden scens that are occluded and outside the line of sight of the imaging system. We densely scan a visible surface. At each scan point, we extract geometric features, which we call Fermat paths, from time-of-flight measurements. Then we apply tools in differential geometry to reconstruct the surface of NLOS objects using the collection of Fermat paths at all scan points. (b) We tackle the problem of reconstructing specular objects. We capture images using a translating camera and a near-field point light source. We first apply the theory of Fermat paths for an initial reconstruction, and then refine it with radiometric measurements of specularities. (c) A commercial off-the-shelf dual-pixel (DP) camera can be considered as a differential stereo system with a tiny baseline. In the presence of defocus blur in the captured DP image, we propose a method to jointly estimate the defocus map and the all-in-focus image using a single DP image.

Chapter 2

A Theory of Fermat Paths

In this chapter, we introduce the theory of *Fermat paths*, which are light paths in the scene that satisfy Fermat's principle. Fermat's principle, proposed by the French mathematician Pierre de Fermat in 1662, is a fundamental principle in geometric optics. The goal of this chapter is to formally define Fermat paths and prove their properties through differential analysis. Most of the discussion in this chapter is based on a so-called confocal setting, but they have straightforward generalization to a non-confocal one.

In Section 2.1, we introduce notations and scene settings. In Section 2.2, we begin our analysis on signed distance functions (SDFs), which is closely related to the shortest path—a special type of Fermat paths. In Section 2.3, we extend the shortest path to general Fermat paths. In Section 2.4, we propose surface reconstruction algorithms using Fermat paths based on their properties. Finally, in Section 2.5, we generalize the theory and algorithms of Fermat paths from a confocal setting to a non-confocal one.

2.1 Problem Setting

We assume a piecewise smooth oriented 2D manifold $\mathcal{X}$ is embedded in the 3D Euclidean space $\mathbb{R}^3$, formed as the union of smooth surfaces that are at least twice differentiable. For any point $\boldsymbol{x} \in \mathcal{X}$ with a well-defined surface normal, we use $\hat{\boldsymbol{n}}(\boldsymbol{x})$ to denote its outward unit normal direction. We use $\partial\mathcal{X} \subset \mathcal{X}$ to denote the set of points on $\mathcal{X}$ where a surface normal is not defined. We will be referring to $\partial\mathcal{X}$ as the *boundary* of $\mathcal{X}$ for simplicity, but note that, in addition to boundary points, it also includes points at discontinuous intersections of the smooth surfaces that make up $\mathcal{X}$.

We explore the space of paths that light can travel from a point A to a point B undergoing one reflection event (bounce) on the surface $\mathcal{X}$. Specifically, we consider two setups: the confocal setup where A and B are co-located and the non-confocal one where the two points are at distinctive locations. We devote most of our discussion on the confocal case for a simpler presentation, but our theory extends to the non-confocal setup as well, which we will present at the end of this chapter.

In the confocal case, we denote a single-bounce light path originates and ends at the same point $\boldsymbol{v}$ as

$$\boldsymbol{p}\left(\boldsymbol{x};\boldsymbol{v}\right) \triangleq \boldsymbol{v} \rightarrow \boldsymbol{x} \rightarrow \boldsymbol{v}\,, \tag{2.1}$$

and its pathlength as

$$\tau\left(\boldsymbol{x};\boldsymbol{v}\right) \triangleq \left\|\boldsymbol{x}-\boldsymbol{v}\right\| + \left\|\boldsymbol{x}-\boldsymbol{v}\right\| = 2\left\|\boldsymbol{x}-\boldsymbol{v}\right\|\,. \tag{2.2}$$

2.2 Signed Distance Function

Within the infinite space of potential paths that light can follow $\{\boldsymbol{p}\left(\boldsymbol{x};\boldsymbol{v}\right), \forall \boldsymbol{x} \in \mathcal{X}\}$, the shortest one is often of interest because it is a strong geometric representation of $\mathcal{X}$. The shortest path is one of the Fermat paths as we will see soon in the next section. In fact, it is considered as *the* Fermat path because Fermat's principle is often *incorrectly* interpreted as the "shortest path principle". Although the statement is not entirely true, it provides an instructive special case that we are very familiar with.

From Equation (2.1), we know that for any point $\boldsymbol{v}$, the shortest round-trip pathlength exactly equals twice the SDF: $\tau_{\text{min}}\left(\boldsymbol{x};\boldsymbol{v}\right) = 2\mathcal{D}\left(\boldsymbol{v}\right)$, where $\mathcal{D}\left(\boldsymbol{v}\right)$ denotes the SDF defined as follows.

We start by defining the nearest set of $\boldsymbol{v}$.

Definition 1. The nearest set. *For any point $\boldsymbol{v}$, the* nearest set $\mathcal{N}\left(\boldsymbol{v}\right) \subset \mathcal{X}$ *consists of points $\boldsymbol{x} \in \mathcal{X}$ that have the shortest distance to $\boldsymbol{v}$:*

$$\mathcal{N}\left(\boldsymbol{v}\right) = \left\{\boldsymbol{x} : \underset{x\in\mathcal{X}}{\text{argmin}} \left\|\boldsymbol{x}-\boldsymbol{v}\right\|\right\}\,. \tag{2.3}$$

We can then define the SDF as follows.

Definition 2. Signed distance function (SDF). *Given an oriented surface $\mathcal{X}$, an SDF $\mathcal{D}$:*

$\mathbb{R}^3 \to \mathbb{R}$ *is a scalar field by mapping* $\boldsymbol{v}$ *to the signed distance to points in its nearest set:*

$$\mathcal{D}(\boldsymbol{v}) \triangleq \begin{cases} \|\boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v}\|, \boldsymbol{x}_{\mathcal{N}} \in \mathcal{N}(\boldsymbol{v}), & \text{if } \boldsymbol{v} \in \mathcal{X}^+, \\ -\|\boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v}\|, \boldsymbol{x}_{\mathcal{N}} \in \mathcal{N}(\boldsymbol{v}), & \text{if } \boldsymbol{v} \in \mathcal{X}^-, \\ 0, & \text{if } \boldsymbol{v} \in \mathcal{X}. \end{cases} \tag{2.4}$$

The sign is determined by whether $\boldsymbol{v}$ *is in the positive* ($\mathcal{X}^+$) *or the negative side of* ($\mathcal{X}^-$) *of the oriented surface* $\mathcal{X}$.

Definition 2 shows that $\mathcal{D}(\boldsymbol{v}) = 0$ is an implict representation of the surface, which is often referred to as the zero set in literatures. An SDF encodes rich geometric information and various analysis can be done using its properties. Next, we use tools from differential geometry to conduct first- and second-order differential analyses to obtain gradient and curvature information of an SDF. In all analyses, we assume that $\boldsymbol{v} \in \mathcal{X}^+$ is on the positive side of the surface.

Note on differentiablity. An SDF is differentiable *almost* everywhere, i.e., depending on $\mathcal{X}$, there could exist a sparse set of points $\boldsymbol{v}$ where SDF is non-differentiable. These special points are on the *skeleton* (or medial axis) of the surface [124], whose nearest set $\mathcal{N}(\boldsymbol{v})$ contains more than one point. For an extreme example, let us consider the center of a concave hemisphere being equidistant to all points on the surface. The following analyses are performed only on the points that are differentiable.

2.2.1 First-order Analysis—the Eikonal Equation

In first-order analysis, we evaluate the spatial gradient of the SDF.

Proposition 3. The eikonal equation. *Consider the SDF* $\mathcal{D}(\boldsymbol{v})$ *evaluated at* $\boldsymbol{v}$. *Assume that there is a unique point* $\boldsymbol{x}_{\mathcal{N}} \in \mathcal{N}(\boldsymbol{v})$. *Then, the spatial gradient* $\hat{\boldsymbol{w}} : \mathbb{R}^3 \to \mathbb{R}^3$ *is a unit vector field*

$$\hat{\boldsymbol{w}}(\boldsymbol{v}) = \nabla_{\boldsymbol{v}} \mathcal{D}(\boldsymbol{v}) = -\frac{\boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v}}{\|\boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v}\|}, \tag{2.5}$$

and naturally

$$\|\hat{\boldsymbol{w}}(\boldsymbol{v})\| = 1. \tag{2.6}$$

Proof. We use $\boldsymbol{v} = [v^x, v^y, v^z]^T$ to denote the 3D coordinates of the point $\boldsymbol{v}$, and similarly for all other vectors. We consider each coordinate of the vector $\nabla_{\boldsymbol{v}} \mathcal{D}(\boldsymbol{v})$ separately.

Differentiating $\mathcal{D}(\boldsymbol{v})$ with respect to v^x yields:

$$\frac{\partial \mathcal{D}(\boldsymbol{v})}{\partial v^x} = \frac{\partial \|\boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v}\|}{\partial v^x} \tag{2.7}$$

$$= \left\langle \frac{\boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v}}{\|\boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v}\|}, \frac{\partial (\boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v})}{\partial v^x} \right\rangle \tag{2.8}$$

$$= \left\langle \frac{\boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v}}{\|\boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v}\|}, \frac{\partial \boldsymbol{x}_{\mathcal{N}}}{\partial v^x} - [1,0,0]^T \right\rangle \tag{2.9}$$

$$= \frac{1}{\|\boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v}\|} \left(\left\langle \boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v}, \frac{\partial \boldsymbol{x}_{\mathcal{N}}}{\partial v^x} \right\rangle - \left\langle \boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v}, [1,0,0]^T \right\rangle \right) . \tag{2.10}$$

Because $\boldsymbol{x}_{\mathcal{N}}$ lies on $\mathcal{X}$, any derivative of $\boldsymbol{x}_{\mathcal{N}}$ is tangent to the surface $\mathcal{X}$ at $\boldsymbol{x}_{\mathcal{N}}$. Additionally, given that we assume $\boldsymbol{x}_{\mathcal{N}} \in \mathcal{N}(\boldsymbol{v})$, we know the vector from $\boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v}$ is normal to $\mathcal{X}$ at $\boldsymbol{x}_{\mathcal{N}}$, i.e., parallel to the surface normal $\hat{\boldsymbol{n}}(\boldsymbol{x}_{\mathcal{N}})$, therefore $\left\langle \boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v}, \frac{\partial \boldsymbol{x}_{\mathcal{N}}}{\partial v^x} \right\rangle = 0$ and Equation (2.10) becomes

$$\frac{\partial \mathcal{D}(\boldsymbol{v})}{\partial v^x} = \frac{1}{\|\boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v}\|} \left(0 - \left\langle \boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v}, [1,0,0]^T \right\rangle \right) = -\frac{(\boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v})^x}{\|\boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v}\|} . \tag{2.11}$$

Exactly analogously, we can prove that

$$\frac{\partial \mathcal{D}(\boldsymbol{v})}{\partial v^y} = -\frac{(\boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v})^y}{\|\boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v}\|} \qquad \text{and} \qquad \frac{\partial \mathcal{D}(\boldsymbol{v})}{\partial v^z} = -\frac{(\boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v})^z}{\|\boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v}\|} . \tag{2.12}$$

Considering all three coordinates yields

$$\hat{\boldsymbol{w}}(\boldsymbol{v}) = \nabla_{\boldsymbol{v}} \mathcal{D}(\boldsymbol{v}) = -\frac{\boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v}}{\|\boldsymbol{x}_{\mathcal{N}} - \boldsymbol{v}\|} , \tag{2.13}$$

and

$$\|\hat{\boldsymbol{w}}(\boldsymbol{v})\| = 1 . \tag{2.14}$$

□

Proposition 3 is the well-known eikonal equation for SDFs [131, 159, 160]. The uniqueness requirement guarantees diffentiability. When the uniqueness condition is violated, the spatial gradient simply does not exist. Proposition 3 reveals two important first-order properties of an SDF. First, the spatial gradient has a unit normal wherever differentiable. Second, the gradient points toward the nearest surface point. Intuitively, if we describe an SDF as a propagating wavefront from the surface, then these two properties indicate that the wavefront is propagating at a constant unit speed along its normal direction.

2.2.2 Second-order Analysis

We can further obtain curvature information of an SDF through second-order analysis by evaluating the Hessian matrix of SDF $\mathcal{H}_{\mathcal{D}}$, or equivalently the Jacobian of its gradient $\mathcal{J}_{\hat{\boldsymbol{w}}}$:

$$\mathcal{H}_{\mathcal{D}}(\boldsymbol{v}) = \mathcal{J}_{\hat{\boldsymbol{w}}}(\boldsymbol{v}) = \begin{bmatrix} \frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^x \partial v^x} & \frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^x \partial v^y} & \frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^x \partial v^z} \\ \frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^x \partial v^y} & \frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^y \partial v^y} & \frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^y \partial v^z} \\ \frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^x \partial v^z} & \frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^y \partial v^z} & \frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^z \partial v^z} \end{bmatrix}. \tag{2.15}$$

Proposition 4. Eigen decomposition of the Hessian matrix. *At a point $\boldsymbol{v}$, the eigen decomposition of the Hessian of an SDF $\mathcal{D}(\boldsymbol{v})$ has the following form*

$$\mathcal{H}_{\mathcal{D}}(\boldsymbol{v}) = \begin{bmatrix} \hat{\boldsymbol{w}}(\boldsymbol{v}) & \hat{\boldsymbol{t}}_{\min}(\boldsymbol{v}) & \hat{\boldsymbol{t}}_{\max}(\boldsymbol{v}) \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & -\kappa_{\min}(\boldsymbol{v}) & 0 \\ 0 & 0 & -\kappa_{\max}(\boldsymbol{v}) \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{w}}(\boldsymbol{v}) \\ \hat{\boldsymbol{t}}_{\min}(\boldsymbol{v}) \\ \hat{\boldsymbol{t}}_{\max}(\boldsymbol{v}) \end{bmatrix}, \tag{2.16}$$

where $\kappa_{\min}(\boldsymbol{v})$ and $\kappa_{\max}(\boldsymbol{v})$ are the principal curvatures of $\mathcal{D}(\boldsymbol{v})$, and $\hat{\boldsymbol{t}}_{\min}(\boldsymbol{v})$ and $\hat{\boldsymbol{t}}_{\max}(\boldsymbol{v})$ the corresponding principal directions.

Proof. From Proposition 3, we have $\|\hat{\boldsymbol{w}}(\boldsymbol{v})\|^2 = 1$. Differentiating both sides with respect to $\boldsymbol{v}$, we have

$$\mathcal{J}_{\hat{\boldsymbol{w}}}(\boldsymbol{v}) \cdot \hat{\boldsymbol{w}}(\boldsymbol{v}) = \mathcal{H}_{\mathcal{D}}(\boldsymbol{v}) \cdot \hat{\boldsymbol{w}}(\boldsymbol{v}) = \mathbf{0}, \tag{2.17}$$

i.e., $\hat{\boldsymbol{w}}$ is in the null space of the Hessian $\mathcal{H}_{\mathcal{D}}$.

Equation (2.17) implies that rows of $\mathcal{H}_{\mathcal{D}}(\boldsymbol{v})$ are orthogonal to $\hat{\boldsymbol{w}}(\boldsymbol{v})$, therefore are all tangent vectors to the level set at $\mathcal{D}(\boldsymbol{v})$. Among all tangent vectors, two of them are along the principal directions $\hat{\boldsymbol{t}}_{\min}(\boldsymbol{v})$ and $\hat{\boldsymbol{t}}_{\max}(\boldsymbol{v})$, corresponding to the principal (smallest and largest) curvatures, leading to the above eigen decomposition. □

Given the Hessian matrix, one can also compute the Gaussian K and mean H curvatures of the SDF as follows:

$$K = \kappa_{\min}\kappa_{\max} = \begin{vmatrix} \frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^x \partial v^x} & \frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^x \partial v^y} \\ \frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^x \partial v^y} & \frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^y \partial v^y} \end{vmatrix} + \begin{vmatrix} \frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^x \partial v^x} & \frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^x \partial v^z} \\ \frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^x \partial v^z} & \frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^z \partial v^z} \end{vmatrix} + \begin{vmatrix} \frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^y \partial v^y} & \frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^y \partial v^z} \\ \frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^y \partial v^z} & \frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^z \partial v^z} \end{vmatrix}, \tag{2.18}$$

$$H = \frac{1}{2}(\kappa_{\min} + \kappa_{\max}) = -\frac{1}{2}\left(\frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^x \partial v^x} + \frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^y \partial v^y} + \frac{\partial^2 \mathcal{D}(\boldsymbol{v})}{\partial v^z \partial v^z}\right). \tag{2.19}$$

2.3 Fermat Paths

Fermat paths is a superset of the shoretest paths. We define Fermat paths as paths that satisfy Fermat's principle, which characterizes paths of *stationary* length with respect to their *local* variations. Obviously, the shortest path, as a global minimum, is a Fermat path. But there could be more: The local stationarity indicates that any paths that are locally shortest, locally longest, or saddle-point paths (locally longest in one direction and locally shortest in the crossing directions) are also Fermat paths. We note that the naming of *Fermat paths* is given by following classical proposition of geometric optics [38, 81, 122, 168].

Fermat paths follow either the law of specular reflection or reflect at specific points at the object's boundary. We define the corresponding surface points $\boldsymbol{x} \in \mathcal{X}$ generating these two categories of Fermat paths as follows.

Definition 5. the Fermat set. *For any point $\boldsymbol{v}$:*

- *The* specular set $\mathcal{S}(\boldsymbol{v}) \subset \mathcal{X}$ *consists of all points $\boldsymbol{x} \in \mathcal{X} \setminus \partial\mathcal{X}$ such that the vector $\boldsymbol{v} - \boldsymbol{x}$ is orthogonal to the tangent plane $T_{\boldsymbol{x}}\mathcal{X}$ of $\mathcal{X}$ at $\boldsymbol{x}$; or equivalently, the vector $\boldsymbol{v} - \boldsymbol{x}$ is parallel to the surface normal $\hat{\boldsymbol{n}}(\boldsymbol{x})$.*
- *The* boundary set $\mathcal{B}(\boldsymbol{v}) \subset \partial\mathcal{X}$ *consists of all points $\boldsymbol{x} \in \partial\mathcal{X}$ such that the vector $\boldsymbol{v} - \boldsymbol{x}$ is orthogonal to the tangent vector $\hat{\boldsymbol{t}}(\boldsymbol{x})$ of $\partial\mathcal{X}$ at $\boldsymbol{x}$.*
- *The* Fermat set $\mathcal{F}(\boldsymbol{v}) \subset \mathcal{X}$ *is the union of these two sets, $\mathcal{F}(\boldsymbol{v}) \triangleq \mathcal{S}(\boldsymbol{v}) \cup \mathcal{B}(\boldsymbol{v})$.*

$\mathcal{S}(\boldsymbol{v})$ and $\mathcal{B}(\boldsymbol{v})$ can also be formally specified using *Fermat's principle* as follows.

Proposition 6. Fermat's principle. *Let $(p, q) \in [0,1]^2$ be a parameterization of the surface $\mathcal{X}$. Then, for any point $\boldsymbol{v}$,*

$$\mathcal{S}(\boldsymbol{v}) = \left\{ \boldsymbol{x} \in \mathcal{X} : \nabla_{(p,q)} \tau(\boldsymbol{x}(p,q); \boldsymbol{v}) = \boldsymbol{0} \right\}. \tag{2.20}$$

Let $r \in [0,1]$ be a parameterization of the surface boundary $\partial\mathcal{X}$. Then, for any point $\boldsymbol{v}$,

$$\mathcal{B}(\boldsymbol{v}) = \{ \boldsymbol{x} \in \partial\mathcal{X} : \partial\tau(\boldsymbol{x}(r); \boldsymbol{v}) / \partial r = 0 \}. \tag{2.21}$$

We provide a proof in Appendix A.1. Proposition 6 is the formal defintion as Fermat's principle which characterizes paths of stationary length with respect to their local variations.

We now define the Fermat pathlength function (FPF) $\tau_{\mathcal{F}}(\boldsymbol{v})$ corresponding to the Fermat set $\mathcal{F}(\boldsymbol{v})$.

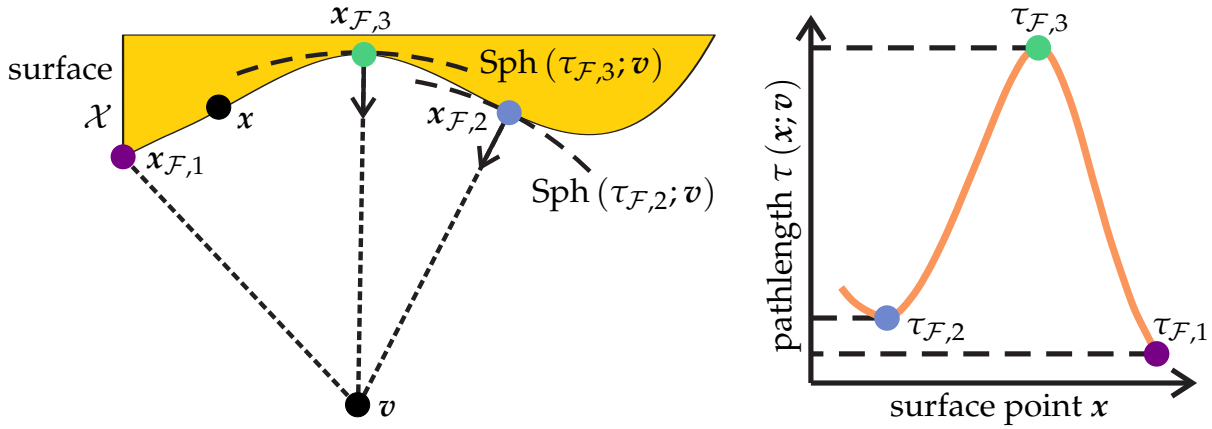


Figure 2.1: **Theory of Fermat paths.** We consider light paths starting from and ending at point v to surface $\mathcal{X}$. Among all points $x \in \mathcal{X}$, some points on the surface ($x_{\mathcal{F},2}$,$x_{\mathcal{F},3}$) and boundary ($x_{\mathcal{F},1}$) will create paths that satisfy Fermat's principle, corresponding to local minima ($x_{\mathcal{F},1}$,$x_{\mathcal{F},2}$) or maxima ($x_{\mathcal{F},3}$) of the pathlength function $\tau(x; v)$ (right). The paths for the non-boundary points ($x_{\mathcal{F},2}$,$x_{\mathcal{F},3}$) will additionally be specular.

Definition 7. Fermat pathlength function (FPF) $\tau_{\mathcal{F}}(v)$**.**

$$\tau_{\mathcal{F}}(v) = \{2\|x_{\mathcal{F}} - v\| : x_{\mathcal{F}} \in \mathcal{F}(v)\}. \tag{2.22}$$

We note the similarity between the FPF $\tau_{\mathcal{F}}(v)$ and the SDF (Definition 2). In fact, SDF (multiplied by 2) is one branch of the FPF because the nearest set is a subset of the Fermat set, i.e., $\mathcal{N}(v) \subset \mathcal{F}(v)$. When $\mathcal{F}(v)$ contains more than one points, as shown in Figure 2.1, FPF becomes a *multi-valued* function. In this case, the differential analysis should be done individually on each branch of the FPF.

The good news is that the analysis on all branches can be performed exactly the same as we did on the shortest path (SDF). In the following subsections, we generalize the differential analysis on SDF in the previous section to FPF, leaving their proof in the appendix.

2.3.1 First-order Analysis

We refer to the following first-order analysis on an FPF as the Fermat flow equation.

Proposition 8. The Fermat flow equation. *Consider a branch of the Fermat pathlength function* $\tau_{\mathcal{F}}(v)$ *evaluated at* v*. Assume that there is a unique point* $x_{\mathcal{F}} \in \mathcal{F}(v)$ *with* $\tau(x_{\mathcal{F}}; v) =$

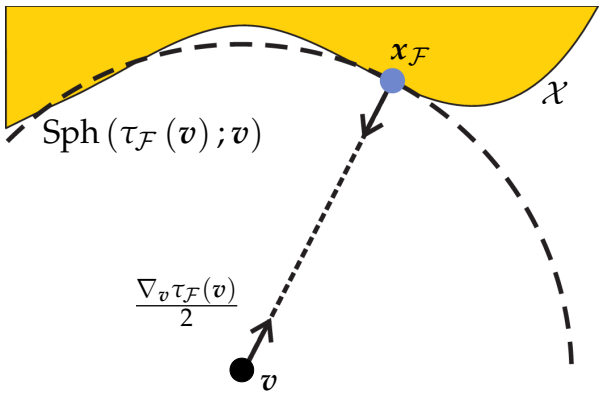


Figure 2.2: **The Fermat flow equation.** We consider the Fermat path connecting the point v with a surface point $x_{\mathcal{F}} \in \mathcal{F}\left(v\right)$. The spatial gradient $\nabla_v \tau_{\mathcal{F}}\left(v\right)/2$ of the length of this path is a unit vector parallel to the vector $x_{\mathcal{F}} - v$. If the Fermat path is also specular, then $\nabla_v \tau_{\mathcal{F}}\left(v\right)/2$ will additionally be the surface normal at $x_{\mathcal{F}}$.

$\tau_{\mathcal{F}}\left(v\right)$. *Then,*

$$\nabla_v \tau_{\mathcal{F}}\left(v\right) = -2\frac{x_{\mathcal{F}} - v}{\left\|x_{\mathcal{F}} - v\right\|} = -2\hat{w}_{\mathcal{F}}\left(v\right) . \tag{2.23}$$

We provide the proof in Appendix A.2. Proposition 8 generalizes the eikonal equation (Proposition 3) to apply to all branches of $\tau_{\mathcal{F}}\left(v\right)$, corresponding to all stationary points $\mathcal{F}\left(v\right)$ of $\tau\left(x;v\right)$ and not only those in the nearest set $\mathcal{N}\left(v\right)$ for the shortest path. The scalar 2 is due to Fermat path being a round-trip path between $x_{\mathcal{F}}$ and v, in contrast to the SDF $\mathcal{D}\left(v\right)$ of a single-trip. This is shown in Figure 2.2.

A lemma of Proposition 8, which we call the integral form of the Fermat flow equation, can then be given by integrating the directions of Fermat paths.

Proposition 9. Integral form of Fermat flow equation. *Consider a branch of unit directions of Fermat paths $\hat{w}_{\mathcal{F}}\left(v\right)$ evaluated at v. Assume that there is a unique point $x_{\mathcal{F}} \in \mathcal{F}\left(v\right)$ with $\tau\left(x_{\mathcal{F}};v\right) = \tau_{\mathcal{F}}\left(v\right)$. Then,*

$$\tau_{\mathcal{F}}\left(v\right) = \int_v -2\hat{w}_{\mathcal{F}}\left(v\right)\,\mathrm{d}v + \tau_0\,, \tag{2.24}$$

where τ_0 denotes an unknown constant.

Proposition 9 is the inverse of Proposition 8. However, unlike differentiation, the integration comes with an unknown constant, i.e., the integrated FPFs is a one-parameter family (OPF). Proposition 9 is essentially the normal integration technique [76, 145, 187] applied to individual branches of the directions of Fermat paths.

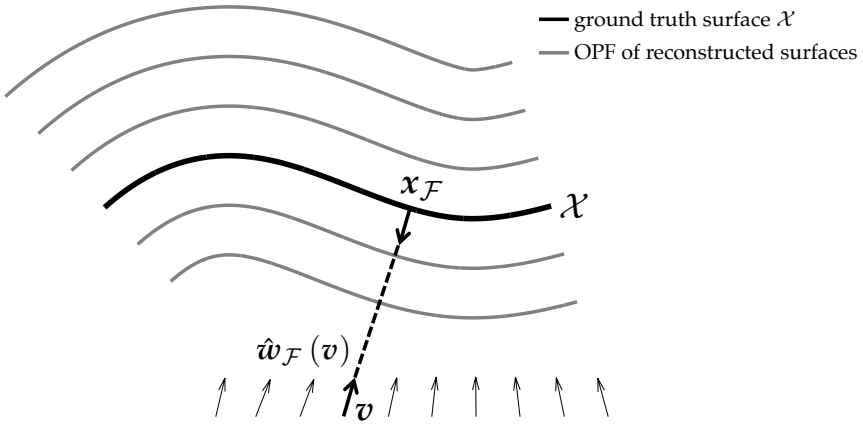


Figure 2.3: **Integral form of Fermat flow equation.** We consider a branch of unit directions of Fermat paths $\hat{w}_{\mathcal{F}}(v)$. We can integrate them to obtain the FPF $\tau(v)$ using the integral form of Fermat flow equation (Proposition 9). The unknown constant in the integral leads to a one-parameter family of FPFs.

2.3.2 Second-order Analysis

We can further conduct second-order analysis on the FPF. Because computing the Hessian matrix requires twice differentiability, the analysis is only applicable to specular paths and not boundary paths because the curvature of a boundary point is not well-defined. We denote the Hessian of one branch of the FPF $\mathcal{H}_{\tau_{\mathcal{S}}}$, or equivalently the Jacobian of its gradient $\mathcal{J}_{\hat{w}_{\mathcal{S}}}$ as

$$\mathcal{H}_{\tau_{\mathcal{S}}}(v) = 2\mathcal{J}_{\hat{w}_{\mathcal{S}}}(v) = 2\begin{bmatrix} \frac{\partial^2 \tau_{\mathcal{S}}(v)}{\partial v^x \partial v^x} & \frac{\partial^2 \tau_{\mathcal{S}}(v)}{\partial v^x \partial v^y} & \frac{\partial^2 \tau_{\mathcal{S}}(v)}{\partial v^x \partial v^z} \\ \frac{\partial^2 \tau_{\mathcal{S}}(v)}{\partial v^x \partial v^y} & \frac{\partial^2 \tau_{\mathcal{S}}(v)}{\partial v^y \partial v^y} & \frac{\partial^2 \tau_{\mathcal{S}}(v)}{\partial v^y \partial v^z} \\ \frac{\partial^2 \tau_{\mathcal{S}}(v)}{\partial v^x \partial v^z} & \frac{\partial^2 \tau_{\mathcal{S}}(v)}{\partial v^y \partial v^z} & \frac{\partial^2 \tau_{\mathcal{S}}(v)}{\partial v^z \partial v^z} \end{bmatrix}, \tag{2.25}$$

then the analysis can be done exactly as in Section 2.2.2.

In addition, because Fermat paths contain not only the globally shortest path, but also many other locally stationary paths, the stationarity type of *specular* paths already provides us with some "weaker" local curvature information about $\mathcal{X}$. Therefore, instead of a quantitative curvature computation, it is also possible to evaluate the local convexity/concavity of the surface $\mathcal{X}$ at $x_{\mathcal{S}}$, qualitatively.

Proposition 10. *Let a point $x_{\mathcal{S}} \in \mathcal{S}(v)$ belong to the specular set. If $\kappa_{min}, \kappa_{max}$ are the principal curvatures of $\mathcal{X}$ at $x_{\mathcal{S}}$, then:*

- *If $\tau_{\mathcal{S}}$ is a local minimum of $\tau(x; v)$, $2/\tau_{\mathcal{S}} < \kappa_{min}$.*

- *If $\tau_{\mathcal{S}}$ is a local maximum of $\tau(\boldsymbol{x};\boldsymbol{v})$, $\kappa_{max} < 2/\tau_{\mathcal{S}}$.*
- *If $\tau_{\mathcal{S}}$ is a saddle point of $\tau(\boldsymbol{x};\boldsymbol{v})$, $\kappa_{min} \leq 2/\tau_{\mathcal{S}} \leq \kappa_{max}$.*

We provide the proof in Appendix A.3, but we can use Figure 2.1 to provide intuition: The pathlength $\tau_{\mathcal{F},2}$ of point $\boldsymbol{x}_{\mathcal{F},2}$ is a local minimum. All $\mathcal{X}$ points in the neighborhood of $\boldsymbol{x}_{\mathcal{F},2}$ are at a distance from $\boldsymbol{v}$ greater than $\tau_{\mathcal{F},2}$, and therefore outside the tangent sphere $\text{Sph}(\tau_{\mathcal{F},2}/2;\boldsymbol{v})$. This implies that the (minimal, in 3D) principal radius of curvature is greater than $\tau_{\mathcal{F},2}/2$. And conversely for the pathlength $\tau_{\mathcal{F},3}$ of point $\boldsymbol{x}_{\mathcal{F},3}$, which is a local maximum. We note that a *sufficient* condition for $\mathcal{X}$ to produce only locally-minimum specular pathlengths is that $\mathcal{X}$ is convex. However, this is not a *necessary* condition: As explained in Proposition 10, it is possible for $\mathcal{X}$ to contain concavities and still produce only locally-minimum specular pathlengths.

2.4 Surface Reconstruction Using Fermat Paths

Based on the the first- and second-order properties of Fermat paths, we propose reconstruction algorithms based on two types of Fermat paths measurements. We leave the discussion on the employed imaging systems in the next two chapters, and focus on discussing the *algorithms* we will be using once we extract these measurements of Fermat paths.

In Section 2.4.1, we consider using a time-of-flight sensor, which measures the photons' travelling time. Time measurements, multiplied by the speed of light, can be converted to pathlength measurements. We will discuss the usage of a time-of-flight sensor and how we exact pathlength information of Fermat paths in Chapter 3. In Section 2.4.2, we consider using a directional sensor that measures the directions of light rays, e.g., a light field camera or a wavefront sensor. An application will be discussed in Chapter 4.

2.4.1 Reconstruction Using the *Lengths* of Fermat Paths

Given a Fermat pathlength measurement $\tau_{\mathcal{F}}(\boldsymbol{x}_{\mathcal{F}};\boldsymbol{v})$ taken at $\boldsymbol{v}$, we can first associate the point $\boldsymbol{x}_{\mathcal{F}} \in \mathcal{F}(\boldsymbol{v})$ with the sphere $\text{Sph}(\tau_{\mathcal{F}}(\boldsymbol{v})/2;\boldsymbol{v})$ of center $\boldsymbol{v}$ and radius $\tau_{\mathcal{F}}(\boldsymbol{v})/2$. We call this the *tangent sphere*, because Proposition 6 implies that, for $\boldsymbol{x} \in \mathcal{S}(\boldsymbol{v})$ or $\boldsymbol{x} \in \mathcal{B}(\boldsymbol{v})$, the sphere $\text{Sph}(\tau_{\mathcal{F}}(\boldsymbol{v})/2;\boldsymbol{v})$ is tangent to $\mathcal{X}$ or $\partial\mathcal{X}$, respectively, at $\boldsymbol{x}$ [122, 168]. See Figure 2.1 for an illustration.

However, this sphere alone is not sufficient for us to determine the point $\boldsymbol{x}_{\mathcal{F}}$. In the following, we develop a procedure for completely determining the point $\boldsymbol{x}_{\mathcal{F}}$ and its normal and curvature. We show that the Fermat flow equation provides a simple constraint that uniquely determines both the depth and normal of a surface point. This derivative is estimated numerically by fitting a smooth pathlength function to a set of measurements. This procedure will produce an oriented point cloud for the surface $\mathcal{X}$. We then apply a final refinement step that computes a smooth mesh by combining both the depth and normal information [39, 89].

Fermat flow algorithm. Knowing the tangent sphere alone is insufficient—we still do not know where exactly the point is on the sphere. To derive the directional (ray) constraint, we use the Fermat flow equation (2.23) and compute the direction from $\boldsymbol{v}$ to $\boldsymbol{x}_{\mathcal{F}}$ as:

$$\hat{\boldsymbol{w}}_{\mathcal{F}}(\boldsymbol{v}) = \frac{\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}}{\|\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}\|} = -\nabla_{\boldsymbol{v}} \tau_{\mathcal{F}}(\boldsymbol{v})/2\,. \tag{2.26}$$

Together with the sphere $\mathrm{Sph}(\tau_{\mathcal{F}}(\boldsymbol{v})/2; \boldsymbol{v})$, we can reconstruct the point $\boldsymbol{x}_{\mathcal{F}}$ as

$$\boldsymbol{x}_{\mathcal{F}} = \boldsymbol{v} + \frac{\tau_{\mathcal{F}}(\boldsymbol{v})}{2} \cdot \hat{\boldsymbol{w}}_{\mathcal{F}}(\boldsymbol{v}) = \boldsymbol{v} - \frac{\tau_{\mathcal{F}}(\boldsymbol{v})}{4} \nabla_{\boldsymbol{v}} \tau_{\mathcal{F}}(\boldsymbol{v})\,. \tag{2.27}$$

Equation (2.27) states that a surface point $\boldsymbol{x}_{\mathcal{F}} \in \mathcal{F}(\boldsymbol{v})$ generating a Fermat path can be uniquely reconstructed from the corresponding point $\boldsymbol{v}$, the length $\tau_{\mathcal{F}}(\boldsymbol{v})$, and the gradient $\nabla_{\boldsymbol{v}} \tau_{\mathcal{F}}(\boldsymbol{v})$. This reconstruction can be done with a simple geometric operation, by intersecting the sphere $\mathrm{Sph}(\tau_{\mathcal{F}}(\boldsymbol{v})/2; \boldsymbol{v})$ with the line $\boldsymbol{v} - \lambda \nabla_{\boldsymbol{v}} \tau_{\mathcal{F}}(\boldsymbol{v})/2$. If the Fermat path is also specular, $\boldsymbol{x}_{\mathcal{F}} \in \mathcal{S}(\boldsymbol{v})$, then we can also reconstruct the normal at $\boldsymbol{x}_{\mathcal{F}}$ as $\hat{\boldsymbol{n}}(\boldsymbol{x}_{\mathcal{F}}) = \nabla_{\boldsymbol{v}} \tau_{\mathcal{F}}(\boldsymbol{v})/2$. This is shown in Figure 2.2.

In addition to reconstructing points and normals, we can also recover the curvature of $\mathcal{X}$ at $\boldsymbol{x}_{\mathcal{S}}$. This can be done in two steps. First, we differentiate $\hat{\boldsymbol{w}}_{\mathcal{S}}(\boldsymbol{v})$ (Equation (2.15)) to obtain the Hessian matrix, and compute Gaussian and mean curvatures (Equations (2.18) and (2.19)). Second, we "back-propagate" the curvature from $\boldsymbol{v}$ to $\boldsymbol{x}_{\mathcal{F}}$. Curvature propagation can be cast as a wavefront tracing problem. Similar to an SDF, we can also consider a branch of FPF as a propagating wavefront from a local patch of the surface to the point $\boldsymbol{v}$. The wavefront is a *geometric* wavefront, defined as the surface of constant pathlength, and should not be confused with the wave-like behaviors of light, such as diffraction and interference. In 2D, a wavefront with curvature κ evolving by a distance $\tau_{\mathcal{S}}(\boldsymbol{v})/2$ has a new curvature

$$\kappa' = \frac{\kappa}{1 + \frac{\tau_{\mathcal{S}}(\boldsymbol{v})}{2}\kappa}\,. \tag{2.28}$$

Equivalently, back-propagating the wavefront by $\tau_{\mathcal{S}}(\boldsymbol{v})/2$, the curvature becomes

$$\kappa = \frac{\kappa'}{1 - \frac{\tau_{\mathcal{S}}(\boldsymbol{v})}{2}\kappa'}. \tag{2.29}$$

In 3D, by considering the evolution of both principal curvatures $\kappa_{\min}$ and $\kappa_{\max}$, we can relate the curvature of the FPF at $\boldsymbol{v}$ to the surface curvature at point $\boldsymbol{x}_{\mathcal{S}}$, which yields

$$\begin{aligned}
K(\boldsymbol{x}_{\mathcal{S}}) &= \kappa_{\min}(\boldsymbol{x}_{\mathcal{S}})\,\kappa_{\max}(\boldsymbol{x}_{\mathcal{S}}) && (2.30)\\
&= \frac{\kappa_{\min}(\boldsymbol{v})}{1 - \kappa_{\min}(\boldsymbol{v}) \cdot \frac{\tau_{\mathcal{S}}(\boldsymbol{v})}{2}} \cdot \frac{\kappa_{\max}(\boldsymbol{v})}{1 - \kappa_{\max}(\boldsymbol{v}) \cdot \frac{\tau_{\mathcal{S}}(\boldsymbol{v})}{2}} && (2.31)\\
&= \frac{K(\boldsymbol{v})}{1 - H(\boldsymbol{v}) \cdot \tau_{\mathcal{S}}(\boldsymbol{v}) + K(\boldsymbol{v}) \cdot \left(\frac{\tau_{\mathcal{S}}(\boldsymbol{v})}{2}\right)^2}. && (2.32)\\
H(\boldsymbol{x}_{\mathcal{S}}) &= \frac{1}{2}\left(\kappa_{\min}(\boldsymbol{x}_{\mathcal{S}}) + \kappa_{\max}(\boldsymbol{x}_{\mathcal{S}})\right) && (2.33)\\
&= \frac{1}{2}\left(\frac{\kappa_{\min}(\boldsymbol{v})}{1 - \kappa_{\min}(\boldsymbol{v}) \cdot \frac{\tau_{\mathcal{S}}(\boldsymbol{v})}{2}} + \frac{\kappa_{\max}(\boldsymbol{v})}{1 - \kappa_{\max}(\boldsymbol{v}) + \frac{\tau_{\mathcal{S}}(\boldsymbol{v})}{2}}\right). && (2.34)
\end{aligned}$$

Differential imaging. Using Equations (2.27), (2.32), and (2.34) requires knowing the first- and second-order derivatives of the FPF at $\boldsymbol{v}$. Using a time-of-flight sensor, we cannot directly measure the derivatives. Also, measurements taken at an isolated point $\boldsymbol{v}$ do not provide us with derivative information. Therefore, in order to infer gradients, we need to take measurements densely with differential intervals.

Although the spatial gradient $\nabla_{\boldsymbol{v}}\tau_{\mathcal{F}}(\boldsymbol{v})$ is a 3-vector, according to Equation (2.23), it has only 2 degree-of-freedom (its norm has to be 2). Therefore, it is unnecessary to take measurements in three dimensions. Instead, measurements taken on a 2D manifold will suffice. The 2D manifold can be of arbitrary shapes. In fact, in some imaging settings (e.g., non-line-of-sight imaging as we will see in Chapter 3), $\boldsymbol{v}$ is constrained to lie on some (uncontrollable) surface $\mathcal{V}$ in 3D. For simplicity, we discuss here only the case when $\mathcal{V}$ is planar, deferring the general case in Appendix A.4. We adopt a scanning scheme by taking measurements densely on a canonical grid on a planar surface $\boldsymbol{v} \in \mathcal{V}$. Without loss of generality, we further assume $\mathcal{V}$ aligns with the $x - y$ plane of the world coordinate frame.

Given a Fermat pathlength $\tau_{\mathcal{F}}(\boldsymbol{v})$ at a point $\boldsymbol{v}$, we can estimate its partial derivatives $\partial\tau_{\mathcal{F}}/\partial x$ and $\partial\tau_{\mathcal{F}}/\partial y$ by locally interpolating Fermat pathlengths measured at nearby points on the plane. We can then infer the derivative with respect to z by noting that

Equation (2.23) implies that $\|\nabla_v \tau_{\mathcal{F}}(v)\| = 2$,

$$\nabla_v \tau_{\mathcal{F}}(v) = \left(\frac{\partial \tau_{\mathcal{F}}}{\partial x}, \frac{\partial \tau_{\mathcal{F}}}{\partial y}, \sqrt{4 - \left(\frac{\partial \tau_{\mathcal{F}}}{\partial x}\right)^2 - \left(\frac{\partial \tau_{\mathcal{F}}}{\partial y}\right)^2} \right) \Bigg|_v . \tag{2.35}$$

Once we compute the gradient, for specular paths, we can estimate the Hessian matrix based on the fact that the gradient is in the null space of the Hessian (Equation (2.17)):

$$\begin{bmatrix} \frac{\partial^2 \tau_{\mathcal{S}}}{\partial x^2} & \frac{\partial^2 \tau_{\mathcal{S}}}{\partial x \partial y} & \frac{\partial^2 \tau_{\mathcal{S}}}{\partial x \partial z} \\ \frac{\partial^2 \tau_{\mathcal{S}}}{\partial x \partial y} & \frac{\partial^2 \tau_{\mathcal{S}}}{\partial y^2} & \frac{\partial^2 \tau_{\mathcal{S}}}{\partial y \partial z} \\ \frac{\partial^2 \tau_{\mathcal{S}}}{\partial x \partial z} & \frac{\partial^2 \tau_{\mathcal{S}}}{\partial y \partial z} & \frac{\partial^2 \tau_{\mathcal{S}}}{\partial z^2} \end{bmatrix} \Bigg|_v \cdot \begin{bmatrix} \frac{\partial \tau_{\mathcal{S}}}{\partial x} \\ \frac{\partial \tau_{\mathcal{S}}}{\partial y} \\ \frac{\partial \tau_{\mathcal{S}}}{\partial z} \end{bmatrix} \Bigg|_v = \mathbf{0} . \tag{2.36}$$

We can estimate the partial derivatives $\frac{\partial^2 \tau_{\mathcal{S}}}{\partial x^2}$, $\frac{\partial^2 \tau_{\mathcal{S}}}{\partial x \partial y}$, and $\frac{\partial^2 \tau_{\mathcal{S}}}{\partial y^2}$ by differentiating $\tau_{\mathcal{S}}(v)$ twice along x and y directions. Once the top left 2×2 submatrix is estimated, we can compute $\frac{\partial^2 \tau_{\mathcal{S}}}{\partial x \partial z}$ and $\frac{\partial^2 \tau_{\mathcal{S}}}{\partial y \partial z}$ from the first two rows of the matrix, and then plug them in the third row to recover $\frac{\partial^2 \tau_{\mathcal{S}}}{\partial z^2}$. Once the entire Hessian matrix is recovered, we can then compute the Gaussian and mean curvatures at v accordingly.

Note on differentiability. We note that the above estimations fail when the $\tau_{\mathcal{F}}$ is non-differentiable at v, e.g., when the uniqueness condition of Proposition 8 is not satisfied. Additionally, we note that the above estimation procedure needs to be performed separately for each branch of the FPF $\tau_{\mathcal{F}}(v)$; otherwise, differentiation across multiple branches would be meaningless and lead to incorrect reconstruction results.

Surface fitting and optimization. The above procedure produces an *oriented* point cloud, of density comparable to the density of measurements on $\mathcal{V}$. We can then use algorithms that take advantage of normal information to fit a surface representation (e.g., triangular mesh) to the point cloud with increased accuracy [89]. Given such an initial surface reconstruction, in Appendix A.5 we describe an optimization procedure, based on the theory of specular path perturbations [39, 81], that refines the fitted surface to account for possible errors due to inaccurate estimation of the gradients $\nabla_v \tau_{\mathcal{F}}(v)$.

2.4.2 Reconstruction Using the *Directions* of Fermat Paths

Given the directional measurement of a Fermat path measurement $\hat{w}_{\mathcal{F}}(v)$ taken at v, we have a ray constraint, forcing the point $x_{\mathcal{F}}$ to be on line $v + \lambda \hat{w}_{\mathcal{F}}(v)$. However, we do not have the pathelength information indicating how far to go along the line.

Fermat pathlength reconstruction. In order to reconstruct the pathlength, we use the integral form of the Fermat flow equation (2.24). The integration also requires taking measurements densely with a differential interval. Therefore, after obtaining a colletions of directions with scanning, one can apply standard normal integral techniques to recover the FPF. It is worth noting that the integration comes with an unknown constant τ_0. Without resolving the ambiguity, we can reconstruct a OPF of surfaces. Once the ambiguity is resolved, reconstructing points, normals, curvatures, and surface fitting can be performed exactly the same as described in the previous section.

2.5 Generalization to a Non-confocal Setup

So far in this chapter, we have focused on discussing the theory and algorithm developed for the confocal case. There are straightforward generalizations to the non-confocal case where the two end points of a single-bounce light path are at different locations. This section states the analogues of definitions, propositions, and algorithms, for the non-confocal case.

In the non-confocal case, we denote a single-bounce light path originates and ends at points $\boldsymbol{v}_s$ and $\boldsymbol{v}_d$ as

$$\boldsymbol{p}\left(\boldsymbol{x}; \boldsymbol{v}_s, \boldsymbol{v}_d\right) \triangleq \boldsymbol{v}_s \to \boldsymbol{x} \to \boldsymbol{v}_d \,, \tag{2.37}$$

and its pathlength as

$$\tau\left(\boldsymbol{x}; \boldsymbol{v}_s, \boldsymbol{v}_d\right) \triangleq \left\|\boldsymbol{x} - \boldsymbol{v}_s\right\| + \left\|\boldsymbol{x} - \boldsymbol{v}_d\right\| \,. \tag{2.38}$$

We additionally denote by $\boldsymbol{h}\left(\boldsymbol{x}; \boldsymbol{v}_s, \boldsymbol{v}_d\right)$ the *half-vector* corresponding to the directions parallel to $\boldsymbol{v}_s - \boldsymbol{x}$ and $\boldsymbol{v}_d - \boldsymbol{x}$.

Definition 5 in the non-confocal case becomes as follows.

Definition 5′. *For any two points $\boldsymbol{v}_s, \boldsymbol{v}_d$:*

- *The* specular set $\mathcal{S}\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right) \subset \mathcal{X}$ *consists of all points $\boldsymbol{x} \in \mathcal{X} \setminus \partial\mathcal{X}$ such that the half-vector $\boldsymbol{h}\left(\boldsymbol{x}; \boldsymbol{v}_s, \boldsymbol{v}_d\right)$ is orthogonal to the tangent plane $T_{\boldsymbol{x}}\mathcal{X}$ of $\mathcal{X}$ at $\boldsymbol{x}$, or equivalently, the half-vector $\boldsymbol{h}\left(\boldsymbol{x}; \boldsymbol{v}_s, \boldsymbol{v}_d\right)$ is also parallel to the surface normal $\hat{\boldsymbol{n}}\left(\boldsymbol{x}\right)$.*
- *The* boundary set $\mathcal{B}\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right) \subset \partial\mathcal{X}$ *consists of all points $\boldsymbol{x} \in \partial\mathcal{X}$ such that the vector $\boldsymbol{h}\left(\boldsymbol{x}; \boldsymbol{v}_s, \boldsymbol{v}_d\right)$ is orthogonal to the tangent vector $\hat{\boldsymbol{t}}\left(\boldsymbol{x}\right)$ of $\partial\mathcal{X}$ at $\boldsymbol{x}$.*
- *The* Fermat set $\mathcal{F}\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right) \subset \mathcal{X}$ *is the union of these two sets, $\mathcal{F}\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right) \triangleq \mathcal{S}\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right) \cup \mathcal{B}\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)$.*

We state the analogue of Proposition 6 (Fermat's principle) for the non-confocal case.

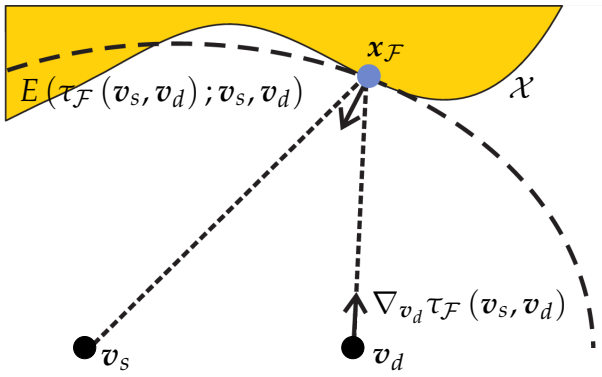


Figure 2.4: **The Fermat flow equation (non-confocal case).** In the non-confocal case, we can compute the gradient with respect to either of the two visible points, v_s or v_d, and it will be parallel to the vector $x_{\mathcal{F}} - v_s$ or $x_{\mathcal{F}} - v_d$, respectively.

Proposition 6′. Fermat's principle. *Let $(p, q) \in [0, 1]^2$ be a parameterization of the surface $\mathcal{X}$. Then, for any pair of points v_s and v_d,*

$$\mathcal{S}(v_s, v_d) = \left\{ x \in \mathcal{X} : \nabla_{(p,q)} \tau(x(p, q); v_s, v_d) = \mathbf{0} \right\}. \tag{2.39}$$

Let $r \in [0, 1]$ be a parameterization of the surface boundary $\partial\mathcal{X}$. Then, for any pair of points v_s and v_d,

$$\mathcal{B}(v_s, v_d) = \{ x \in \partial\mathcal{X} : \partial\tau(x(r); v_s, v_d) / \partial r = 0 \}. \tag{2.40}$$

We provide the proof in Appendix A.6.

We state the analogue of Definition 7 for the non-confocal case.

Definition 7′. Fermat pathlength function (FPF) $\tau_{\mathcal{F}}(v_s, v_d)$**.**

$$\tau_{\mathcal{F}}(v_s, v_d) = \{ \|x_{\mathcal{F}} - v_s\| + \|x_{\mathcal{F}} - v_d\| : x_{\mathcal{F}} \in \mathcal{F}(v_s, v_d) \}. \tag{2.41}$$

2.5.1 First-order Analysis

We state the analogue of Proposition 8 (Fermat flow equation) and Proposition 9 (integral form of Fermat flow equation) for the non-confocal case.

Proposition 8′. Fermat flow equation. *Consider a branch of the Fermat pathlength function $\tau_{\mathcal{F}}(v_s, v_d)$ corresponding to points v_s, v_d. Assume that there is a unique point $x_{\mathcal{F}} \in \mathcal{F}(v_s, v_d)$*

with $\tau\left(\boldsymbol{x}_{\mathcal{F}};\boldsymbol{v}_s,\boldsymbol{v}_d\right)=\tau_{\mathcal{F}}\left(\boldsymbol{v}_s,\boldsymbol{v}_d\right)$*. Then,*

$$\nabla_{\boldsymbol{v}_s}\tau_{\mathcal{F}}\left(\boldsymbol{v}_s,\boldsymbol{v}_d\right)=-\frac{\boldsymbol{x}_{\mathcal{F}}-\boldsymbol{v}_s}{\left\|\boldsymbol{x}_{\mathcal{F}}-\boldsymbol{v}_s\right\|}=-\hat{\boldsymbol{w}}_{\mathcal{F}}\left(\boldsymbol{v}_s,\boldsymbol{v}_d\right)\,,\qquad \textit{and} \tag{2.42}$$

$$\nabla_{\boldsymbol{v}_d}\tau_{\mathcal{F}}\left(\boldsymbol{v}_s,\boldsymbol{v}_d\right)=-\frac{\boldsymbol{x}_{\mathcal{F}}-\boldsymbol{v}_d}{\left\|\boldsymbol{x}_{\mathcal{F}}-\boldsymbol{v}_d\right\|}=-\hat{\boldsymbol{w}}_{\mathcal{F}}\left(\boldsymbol{v}_s,\boldsymbol{v}_d\right)\,. \tag{2.43}$$

We provide the proof in Appendix A.7. We note that, in the proof of Proposition 8′, we effectively use the *orthotomic* to convert the problem into an equivalent problem for a confocal case [26, 27]. Please refer to the appendix for more details.

Proposition 9′. Integral form of Fermat flow equation. *Consider a branch of directions of Fermat paths* $\hat{\boldsymbol{w}}_{\mathcal{F}}\left(\boldsymbol{v}_s,\boldsymbol{v}_d\right)$ *corresponding to points* $\boldsymbol{v}_s,\boldsymbol{v}_d$*. Assume that there is a unique point* $\boldsymbol{x}_{\mathcal{F}}\in\mathcal{F}\left(\boldsymbol{v}_s,\boldsymbol{v}_d\right)$ *with* $\tau\left(\boldsymbol{x}_{\mathcal{F}};\boldsymbol{v}_s,\boldsymbol{v}_d\right)=\tau_{\mathcal{F}}\left(\boldsymbol{v}_s,\boldsymbol{v}_d\right)$*. Then,*

$$\tau_{\mathcal{F}}\left(\boldsymbol{v}_s,\boldsymbol{v}_d\right)=\int_{\boldsymbol{v}_d}-\hat{\boldsymbol{w}}_{\mathcal{F}}\left(\boldsymbol{v}_s,\boldsymbol{v}_d\right)\,\mathrm{d}\boldsymbol{v}_d+\tau_0\,,\qquad \textit{and} \tag{2.44}$$

$$\tau_{\mathcal{F}}\left(\boldsymbol{v}_s,\boldsymbol{v}_d\right)=\int_{\boldsymbol{v}_s}-\hat{\boldsymbol{w}}_{\mathcal{F}}\left(\boldsymbol{v}_s,\boldsymbol{v}_d\right)\,\mathrm{d}\boldsymbol{v}_s+\tau_0\,, \tag{2.45}$$

where τ_0 *represents an unknown constant.*

2.5.2 Second-order Analysis

We state the analogue of Proposition 10 for the non-confocal case.

Proposition 10′. *Let a point* $\boldsymbol{x}_{\mathcal{S}}\in\mathcal{S}\left(\boldsymbol{v}_s,\boldsymbol{v}_d\right)$ *belong to the specular set. Let* $E\left(\tau_{\mathcal{S}};\boldsymbol{v}_s,\boldsymbol{v}_d\right)$ *be the corresponding osculating ellipsoid. If* $\kappa_{min},\kappa_{max}$ *are the principal curvatures of* $\mathcal{X}$ *at* $\boldsymbol{x}_{\mathcal{S}}$*, and* $\lambda_{min},\lambda_{max}$ *the principal curvatures of* $E\left(\tau_{\mathcal{S}};\boldsymbol{v}_s,\boldsymbol{v}_d\right)$ *at* $\boldsymbol{x}_{\mathcal{S}}$*, then:*

- *If* $\tau_{\mathcal{S}}$ *is a local minimum of* $\tau\left(\boldsymbol{x};\boldsymbol{v}_d,\boldsymbol{v}_s\right)$*, then* $\lambda_{min}\leq\kappa_{min}$*.*
- *If* $\tau_{\mathcal{S}}$ *is a local maximum of* $\tau\left(\boldsymbol{x};\boldsymbol{v}_d,\boldsymbol{v}_s\right)$*, then* $\kappa_{max}\leq\lambda_{max}$*.*

We provide the proof in Appendix A.8.

2.5.3 Surface Reconstruction using Fermat Paths

Reconstructing surface points, normals, and curvatures for the non-confocal case follows exactly as in the confocal case (Section 2.4) . The analogue of the tangent sphere is the *osculating ellipsoid* $E\left(\tau_{\mathcal{F}}\left(\boldsymbol{v}_s,\boldsymbol{v}_d\right);\boldsymbol{v}_s,\boldsymbol{v}_d\right)$ of pathlength $\tau_{\mathcal{F}}\left(\boldsymbol{v}_s,\boldsymbol{v}_d\right)$ and foci $\boldsymbol{v}_s$ and $\boldsymbol{v}_d$. Using the Fermat flow equation, we can obtain a ray constraint $\boldsymbol{v}_d+\lambda\hat{\boldsymbol{w}}_{\mathcal{F}}\left(\boldsymbol{v}_s,\boldsymbol{v}_d\right),\lambda\in$

$[0, \infty]$, and reconstruct the point $\boldsymbol{x}$ through an ellipsoid-ray intersection. We can additionally reconstruct the surface normal at the specular point $\hat{\boldsymbol{n}}_{\mathcal{F}}$ as the bisector of the incident and outgoing directions. This is shown in Figure 2.4.

The geometric operation of ellipsoid-ray intersection can also be described algebraically by the incident and outgoing pathlengths given the total Fermat pathlength and the ray direction. We use $\tau_{\mathcal{F}}^{(i)}$ and $\tau_{\mathcal{F}}^{(r)}$ to denote the incident and outgoing pathlengths, respectively,

$$\tau_{\mathcal{F}}^{(i)} = \|\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}_s\| \text{ and } \tau_{\mathcal{F}}^{(r)} = \|\boldsymbol{v}_d - \boldsymbol{x}_{\mathcal{F}}\| . \tag{2.46}$$

We provide the expressions as follows, and refer to Appendix A.9 for the derivations.

$$\tau_{\mathcal{F}}^{(i)}(\boldsymbol{v}_s, \boldsymbol{v}_d) = \frac{{\tau_{\mathcal{F}}}^2 + 2\langle \boldsymbol{v}_d - \boldsymbol{v}_s, \hat{\boldsymbol{w}}_{\mathcal{F}} \rangle \tau_{\mathcal{F}} + \|\boldsymbol{v}_d - \boldsymbol{v}_s\|^2}{2(\tau_{\mathcal{F}} + \langle \boldsymbol{v}_d - \boldsymbol{v}_s, \hat{\boldsymbol{w}}_{\mathcal{F}} \rangle)}, \text{ and} \tag{2.47}$$

$$\tau_{\mathcal{F}}^{(r)}(\boldsymbol{v}_s, \boldsymbol{v}_d) = \frac{{\tau_{\mathcal{F}}}^2 - \|\boldsymbol{v}_d - \boldsymbol{v}_s\|^2}{2(\tau_{\mathcal{F}} + \langle \boldsymbol{v}_d - \boldsymbol{v}_s, \hat{\boldsymbol{w}}_{\mathcal{F}} \rangle)} . \tag{2.48}$$

2.6 Summary

In this chapter, we have introduced a novel theory of Fermat paths. We started our discussion on a special Fermat path—the globally shortest path, which is closely related to the well-known SDF. We presented and proved the important properties of an SDF through differential analysis. We then formally define the Fermat paths and generalize the previous analyses on SDFs to FPFs. Based on the properties of an FPF, we propose reconstruction algorithms based on two types of measurements of Fermat paths, their pathlengths or directions. In order to apply the algorithms in practice, one needs to adopt differential imaging to densely measure information of Fermat paths. These measurements can then be differentiated or integrated for shape reconstruction. We discussed practical considerations of the algorithm, including gradient estimation and surface fitting and optimization. In the last section, we generalize the analyses for the confocal case to a non-confocal one. In the following two chapters, we will apply the theory and algorithms of Fermat paths to solve two concrete real-world reconstruction problems—non-line-of-sight imaging and specular object reconstruction.

Chapter 3

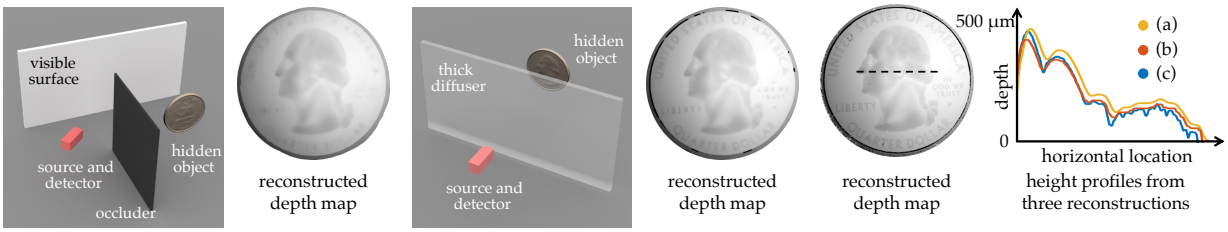
Fermat Paths for Non-line-of-sight Shape Reconstruction

In Chapter 2, we have introducted the theory of Fermat paths and presented surface reconstruction algorithms based on dense measurements of Fermat paths using a differential imaging scheme. In this chapter, we apply the theory to tackle a challenging problem called non-line-of-sight (NLOS) imaging.

3.1 Non-line-of-sight (NLOS) Imaging

Most computer vision research assumes that the scene of interest is directly visible to the camera. In other words, the photons from a source that reach a camera are assumed to have interacted with only the visible scene. However, some of the source photons are reflected by the visible scene toward parts—say, the back of an object facing a camera, an object around a corner, or an object viewed through a diffuser—that are hidden from the direct line of sight of the camera. In turn, the hidden scene scatters the photons back toward the visible scene, which then redirects photons toward the camera. Imaging and understanding the scene hidden from the camera's view is of significant importance to many security and safety applications.

Capturing *non-line-of-sight* (NLOS) photons is challenging as they are vastly outnumbered by line-of-sight (LOS) photons. Passive approaches analyze the subtle umbra and penumbra of the shadow cast by the hidden scene to estimate rough motion and structure [15, 24, 155], or use coherence properties of light to localize hidden ob-



(a) looking around the corner (b) looking through a diffuser (c) comparison with line-of-sight scanning

Figure 3.1: **Non-line-of-sight (NLOS) imaging.** We consider the problem of reconstructing surfaces that are: (a) outside the field of view of sensor, or (b) occluded from it by a diffuser. We develop an algorithm that can use transient imaging measurements to accurately reconstruct the shape of the NLOS surface. The figure shows example reconstructions of a US quarter from measurements captured by a femtosecond-scale transient imaging system. In (c), we compare our reconstructions against groundtruth, obtained using a direct depth scan of the object with the same transient imaging system.

jects [20, 23]. These approaches do not have sufficient information to compute precise 3D shape of an unknown arbitrary hidden scene. Extracting additional information about the hidden scene is possible by using active illumination, including coherent lighting [21, 87, 88, 165] and steady-state intensity sources [93, 171, 174, 189]. The majority of approaches for reconstructing hidden *shape* information employ fast modulated light sources together with time-resolved sensors (e.g., continuous-wave ToF [70, 85], ultrafast photodiodes [92], streak cameras [62, 179, 180], and single-photon avalanche photodetectors (SPADs) [55, 134]). These sensors record not only the number of incident photons (intensity) but also their arrival times, at a range of temporal resolutions (milli- to femto-seconds) [55, 56, 82, 133, 134, 180]. Such measurements are called *transients* and the approach is called *transient NLOS imaging*.

By measuring transients at various locations of a known visible scene, most active techniques perform a volumetric 3D reconstruction by attempting to invert the time-resolved radiometric image formation process. Examples include elliptic backprojection [9, 29, 98, 139, 179], regularized linear system approaches [62, 70, 71, 85], the light-cone transform [135], and analysis-by-synthesis using rendering [138, 176]. These methods have two fundamental disadvantages: (1) they rely on radiometric information and existing SPADs produce poor intensity estimates due to effects such as pile-up and after-pulsing [73], as well as due to extreme sensitivity to photon noise and ambient lighting; and (2) to simplify the inverse problem, all existing reconstruction techniques rely on

an assumption of Lambertian reflectance for the NLOS object.

In this chapter, we overcome the above limitations by developing techniques that use only *geometric*, rather than intensity, constraints derived from transient measurements of an NLOS scene. For this, we apply the theory of Fermat paths (Chapter 2) in the context of NLOS imaging by extracting NLOS photons that follow Fermat paths between the LOS and NLOS scene. We prove that Fermat paths correspond to *discontinuities* in the transient measurements. The temporal locations of the discontinuities are a function of only the shape of the NLOS object and not its reflectance (BRDF). We additionally show that the shape of the transient around the discontinuity is related to the curvature of the hidden surface. The theory of Fermat paths generalizes previous work on the shortest paths of first-returning photons [175], which are a special case of Fermat paths as discussed in Chapter 2.

We then use the Fermat flow algorithm developed in Section 2.4.1 for accurate NLOS shape reconstruction. While most previous approaches reconstruct an albedo volume of the NLOS object, our approach is one of the few that reconstruct its surface. Compared to alternative surface reconstruction algorithms based on analysis-by-synthesis from intensity measurements [176], our approach uses only geometric constraints, which makes it BRDF-invariant and robust to imperfections in intensity measurements.

Our theory is agnostic to the specific transient imaging technology used. We validate our theory and demonstrate results at both pico-second and femto-second temporal scales, using a pulsed laser and SPAD for the former and interferometry for the latter. Hence, for the first time, we are able to compute millimeter-scale and micrometer-scale NLOS shapes of curved objects with BRDFs ranging from purely diffuse to purely specular. In addition, our theory applies to both reflective NLOS (looking around the corner) and transmissive NLOS (seeing through a diffuser) scenarios. Figure 3.1 shows the estimated micrometer-scale relief of a coin seen around the corner as well as through thick paper (diffuser). The obtained height profiles compare well with the reconstruction of the coin when imaged in the *line of sight*. This result demonstrates the significant theoretical and practical contribution of this work to active NLOS imaging, pushing the boundary of what is possible.

3.2 Fermat Paths in NLOS Transients

Imaging setup. We consider a *transient imaging* system [83], comprising a light source and detector, located at points $\boldsymbol{s}, \boldsymbol{d} \in \mathbb{R}^3$, respectively. Our theory is agnostic to the specific transient imaging technology used, and in Section 3.4 we describe implementations, one based on a pulsed laser and a picosecond detector, and another based on interferometry. The *visible* scene $\mathcal{V} \subset \mathbb{R}^3$ is the union of surfaces contained within the common line of sight of the source and detector. In addition to $\mathcal{V}$, we assume that there exist surfaces outside their line of sight; this could be because either these surfaces are outside the field of view, or they are inside it but occluded by another surface. We are only interested in such surfaces that can *indirectly* receive light from the light source by means of a single reflection or transmission through the visible scene, and can *indirectly* send light to the detector in a likewise manner. We call the union of such surfaces the *non-line-of-sight* (NLOS) scene $\mathcal{X}$. Some situations where these conditions apply, and which will be relevant to our experiments, are shown in Figure 3.1.

We assume that the light source and detector are illuminating and imaging the same *visible point* $\boldsymbol{v} \in \mathcal{V}$, which can be any point in the visible scene. This corresponds to the *confocal* scanning scheme, proposed by O'Toole et al. [135]. We emphasize that this assumption is only to simplify exposition: As we have discussed in Section 2.5, all of our theory generalizes to the non-confocal case. In particular, in Section 3.4, we show results from non-confocal experiments.

The detector records a *transient* $I(t; \boldsymbol{v})$, which equals the irradiance from photons with time of flight t. We assume that all recorded photons follow paths of the form $\boldsymbol{s} \to \boldsymbol{v} \to \boldsymbol{x} \to \boldsymbol{v} \to \boldsymbol{d}$, where $\boldsymbol{x} \in \mathcal{X}$. This *three-bounce* assumption is commonplace in NLOS imaging applications, for two reasons: First, NLOS transient imaging systems typically have time-gating mechanisms that can be used to remove *direct* photons that only interact with the visible scene. Second, photons with more than one interactions with the NLOS scene $\mathcal{X}$ have greatly reduced signal-to-noise ratio, and in practice are difficult to detect [29, 138].

Finally, we assume that we have calibrated the distance $\tau_{\mathcal{V}}(\boldsymbol{v}) \triangleq \|\boldsymbol{s} - \boldsymbol{v}\| + \|\boldsymbol{d} - \boldsymbol{v}\|$ from the source to the visible point, and from there to the detector. Then, we can use the pathlength traveled in $\mathcal{X}$, $\tau \triangleq ct - \tau_{\mathcal{V}}(\boldsymbol{v})$ where c is the speed of light, to uniquely

reparameterize transients as $I(\tau; \boldsymbol{v})$. Under these assumptions, we can write [45, 140]:

$$I(\tau; \boldsymbol{v}) = \int_{\mathcal{X}} f(\boldsymbol{x}; \boldsymbol{v}) \, \delta(\tau - \tau(\boldsymbol{x}; \boldsymbol{v})) \, \mathrm{d}A(p, q) \,, \tag{3.1}$$

where $\tau(\boldsymbol{x}; \boldsymbol{v}) \triangleq 2 \cdot \|\boldsymbol{x} - \boldsymbol{v}\|$, $(p, q) \in [0, 1]^2$ is a parameterization of the NLOS surface $\mathcal{X}$, $A(p, q)$ is the corresponding area measure, and the *throughput* f absorbs inverse-square fall-off, shading, reflectance, and visibility. Through this reparameterization, we have effectively converted the NLOS imaging setup to the same setup as in Section 2.1, as if light paths directly starting from and ending at $\boldsymbol{v}$.

3.2.1 Fermat Pathlengths as Transient Discontinuities

Except when the BRDF of the NLOS surface $\mathcal{X}$ is perfectly specular, transients $I(\tau; \boldsymbol{v})$ will include contributions from photons that follow both Fermat and non-Fermat paths $\boldsymbol{p}(\boldsymbol{x}; \boldsymbol{v}) \triangleq \boldsymbol{v} \to \boldsymbol{x} \to \boldsymbol{v}$. Without prior knowledge of the scene, it would seem impossible to identify parts of the transient due to Fermat paths. However, we make the following observation.

Proposition 11. *Assume that the BRDF of the surface $\mathcal{X}$ is non-zero in the specular direction. Then, for all $\boldsymbol{x} \in \mathcal{F}(\boldsymbol{v})$, the transient $I(\tau; \boldsymbol{v})$ will have a discontinuity at pathlength $\tau(\boldsymbol{x}; \boldsymbol{v})$. If $\boldsymbol{x} \in \mathcal{S}(\boldsymbol{v})$, then $I(\tau; \boldsymbol{v})$ will additionally have a vertical asymptote at $\tau(\boldsymbol{x}; \boldsymbol{v})$.*

Proof sketch. We sketch a proof for the specular case, and provide the full proof in Appendix B.1. Let $\mathrm{Sph}(\rho; \boldsymbol{v})$ be the sphere of center $\boldsymbol{v}$ and radius ρ. Let the curve $\mathcal{C}(\rho; \boldsymbol{v})$ be the intersection of $\mathrm{Sph}(\rho; \boldsymbol{v})$ with $\mathcal{X}$, parameterized by $r \in [0, 1]$. Then, we can use $(r, \rho) \in [0, 1] \times [0, \infty)$ to reparameterize $\mathcal{X}$, and rewrite the integral of Equation (3.1):

$$I(\tau; \boldsymbol{v}) = \int_{\mathcal{X}} f(\boldsymbol{x}; \boldsymbol{v}) \, \delta(\tau - \tau(\boldsymbol{x}; \boldsymbol{v})) \left| \mathcal{J}^{(r,\rho)}_{(p,q)}(\boldsymbol{x}) \right|^{-1} \mathrm{d}A(r, \rho) \,, \tag{3.2}$$

where $\mathcal{J}^{(r,\rho)}_{(p,q)}(\boldsymbol{x})$ is the Jacobian of the transformation $(p, q) \mapsto (r, \rho)$. We now consider a point $\boldsymbol{x}_{\mathcal{S}} \in \mathcal{S}(\boldsymbol{v})$. Recognizing that $\rho(\boldsymbol{x}_{\mathcal{S}}) = \tau(\boldsymbol{x}_{\mathcal{S}}; \boldsymbol{v})/2$, we have from Equation (2.20) that $\nabla_{(p,q)} \rho(\boldsymbol{x}_{\mathcal{S}}) = \boldsymbol{0}$. Consequently,

$$\left| \mathcal{J}^{(r,\rho)}_{(p,q)}(\boldsymbol{x}_{\mathcal{S}}) \right| = \frac{\partial \rho(\boldsymbol{x}_{\mathcal{S}})}{\partial p} \frac{\partial r(\boldsymbol{x}_{\mathcal{S}})}{\partial q} - \frac{\partial \rho(\boldsymbol{x}_{\mathcal{S}})}{\partial q} \frac{\partial r(\boldsymbol{x}_{\mathcal{S}})}{\partial p} = 0. \tag{3.3}$$

Then, from Equation (3.2), at $\tau = \tau(\boldsymbol{x}_{\mathcal{S}}; \boldsymbol{v})$, the transient converges to infinity, resulting in a discontinuity. □

Figure 3.2 visualizes this proposition for a *two-dimensional* Lambertian scene $\mathcal{X}$, and a visible point $\boldsymbol{v}$ such that $\mathcal{S}(\boldsymbol{v}) = \{\boldsymbol{x}_{\mathcal{F},2}, \boldsymbol{x}_{\mathcal{F},3}\}$, $\mathcal{B}(\boldsymbol{v}) = \{\boldsymbol{x}_{\mathcal{F},1}\}$. We note that, in two dimensions, the boundary $\partial\mathcal{X}$ is not a curve but just isolated points, and therefore the tangency property of Definition 5 and the tangent sphere are not meaningful.

BRDF invariance. Proposition 11 implies that the pathlengths where the transient $I(\tau; \boldsymbol{v})$ is discontinuous are determined completely by the function $\tau(\boldsymbol{x}; \boldsymbol{v})$. In turn, $\tau(\boldsymbol{x}; \boldsymbol{v})$ depends only on the *geometry* of $\boldsymbol{v}$ and $\mathcal{X}$. Therefore, the discontinuity pathlengths are independent of the BRDF of the NLOS surface $\mathcal{X}$. The BRDF is included in the throughput term f in Equation (3.2), and thus only affects the intensity of the transient at the discontinuity pathlength. Figure 3.3 demonstrates this reflectance invariance property.

Identifying type of stationarity. Proposition 11 allows us to identify the lengths of all Fermat paths that contribute to a transient $I(\tau; \boldsymbol{v})$, as the pathlengths where $I(\tau; \boldsymbol{v})$ is discontinuous. From Proposition 6 in Chapter 2, each of these pathlengths is a stationary point of the function $\tau(\boldsymbol{x}; \boldsymbol{v})$. When the BRDF of $\mathcal{X}$ is not perfectly specular, we can additionally identify the type of stationarity from the shape of the transient at a neighborhood of the discontinuity. We use Figure 3.2 for intuition. Specifically, let $\tau_{\mathcal{F}}$ be a Fermat pathlength where the transient is discontinuous. If $\tau_{\mathcal{F}}$ is a local maximum, the discontinuity in the transient $I(\tau; \boldsymbol{v})$ occurs at the limit from the left, $\tau \to \tau_{\mathcal{F}}^{-}$, and the transient decreases to the right of $\tau_{\mathcal{F}}$ (Figure 3.2, $\tau_{\mathcal{F},3}$). Conversely, when $\tau_{\mathcal{F}}$ is a local minimum, the discontinuity occurs at the limit from the right, $\tau \to \tau_{\mathcal{F}}^{+}$ (Figure 3.2, $\tau_{\mathcal{F},1}$,$\tau_{\mathcal{F},2}$). Finally, when $\tau_{\mathcal{F}}$ is a saddle point, the discontinuity and intensity rise are two-sided. An example of this is shown in Figure 3.3 (paraboloid case).

Identifying the stationarity type of *specular* discontinuities provides us with curvature information about $\mathcal{X}$ (Proposition 10).

Relationship to first-returning photons. Fermat paths are a superset of the paths of the first-returning photons, i.e., the shortest paths, described by Tsai *et al*. [175]. In particular, the pathlength of the first-returning photon is the *global* minimum of $\tau(\boldsymbol{x}; \boldsymbol{v})$. Observations 2 and 3 of Tsai *et al*. [175], which make an assumption of local smoothness, correspond to the case where additionally $\boldsymbol{x} \in \mathcal{S}(\boldsymbol{v})$: Observation 3 describes the specular path $\boldsymbol{p}(\boldsymbol{x}; \boldsymbol{v})$, and Observation 2 describes the tangent sphere.

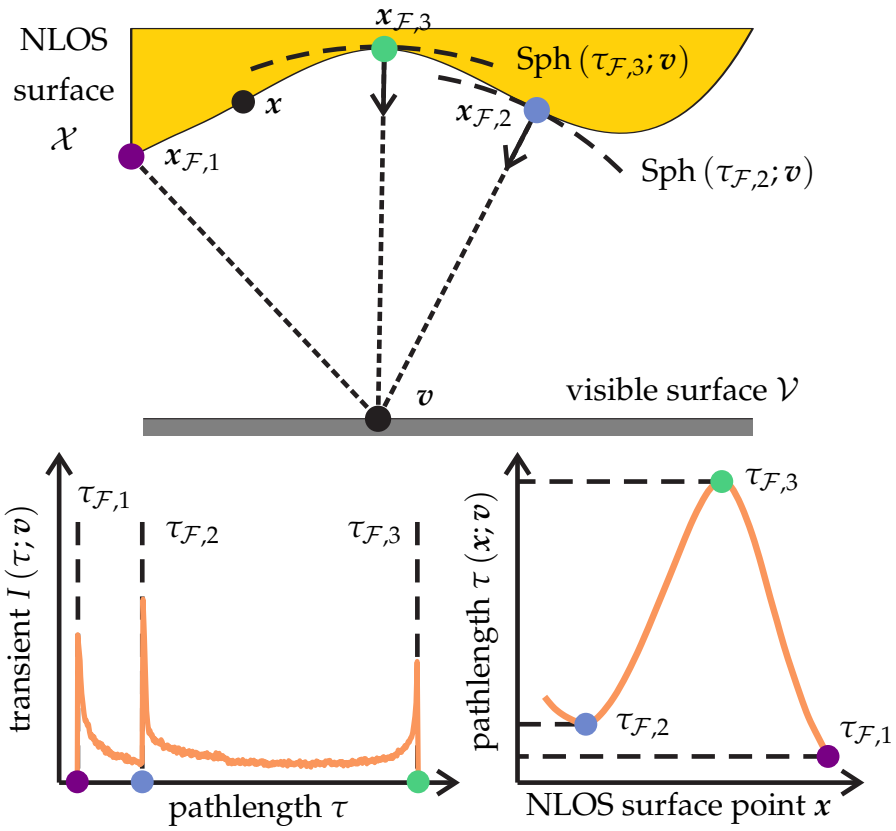


Figure 3.2: **Fermat paths in NLOS imaging.** We illuminate and image an NLOS surface $\mathcal{X}$ from a point v on a visible surface $\mathcal{V}$. (We show the camera and light sources in Figure 3.1.) Among all points $x \in \mathcal{X}$, some points on the surface ($x_{\mathcal{F},2}$,$x_{\mathcal{F},3}$) and boundary ($x_{\mathcal{F},1}$) will create paths that satisfy Fermat's principle, corresponding to local minima ($x_{\mathcal{F},1}$,$x_{\mathcal{F},2}$) or maxima ($x_{\mathcal{F},3}$) of the pathlength function $\tau(x; v)$ (bottom right). The paths for the non-boundary points ($x_{\mathcal{F},2}$,$x_{\mathcal{F},3}$) will additionally be specular. We can identify the lengths of these Fermat paths from the fact that the transient $I(\tau; v)$ (bottom left) will be discontinuous at the corresponding pathlengths ($\tau_{\mathcal{F},1}$,$\tau_{\mathcal{F},2}$,$\tau_{\mathcal{F},3}$).

3.2.2 Experimental Demonstration

To demonstrate our theoretical findings in practice, we use a picosecond-resolution transient imaging setup (see Section 3.4.1) to capture measurements of a few real-world objects, in a looking-around-the-corner configuration (Figure 3.1(a)). Figure 3.3 shows the objects: A concave hemisphere, an extruded paraboloid, and a plane. All objects have size 20 cm $\times$ 20 cm and are painted with diffuse paint.

We place each object at a distance 40 cm from the visible wall, then measure a transient from a visible point such that there is a *specular* path corresponding roughly to the

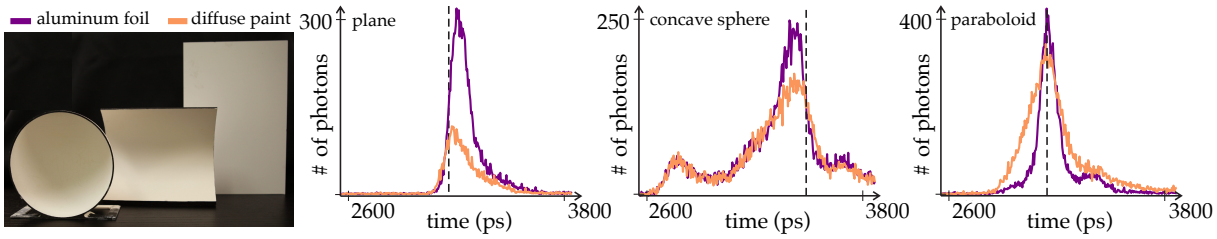


Figure 3.3: **Experimental demonstration.** We measure transients for three objects in a looking-around-the-corner configuration: A plane, a paraboloid, and a concave sphere. We measure each object twice, once with the object covered with diffuse paint, and a second time with the object covered with aluminum foil. As predicted by our theory, the measured transients have discontinuities corresponding to specular paths of type local minimum, saddle point, and local maximum, respectively. Additionally, the location of the discontinuities is not affected by the change in BRDF.

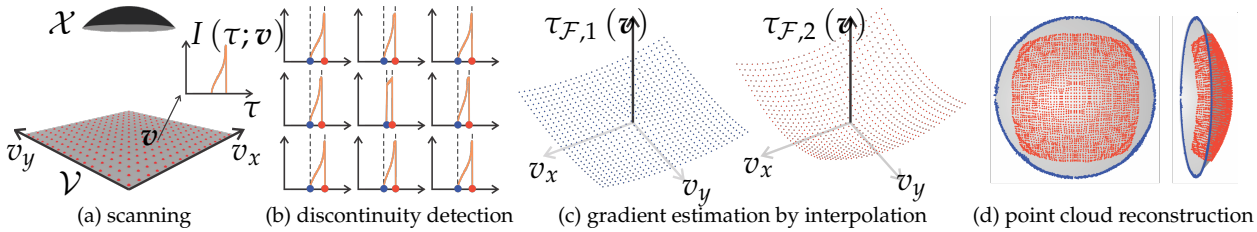
center of the object. We observe from the measured transients (Figure 3.3, orange) that, in agreement with Proposition 10, the hemisphere produces a local maximum discontinuity, the paraboloid a saddle point, and the plane a local minimum.

We then cover each object with aluminum foil, to create a rough specular BRDF, and repeat our measurements. We notice that the measured transients (Figure 3.3, purple) are discontinuous at the same locations as the diffuse transients, in agreement with our discussion of BRDF invariance.

These measurements additionally help evaluate the robustness of our theoretical predictions in the presence of the Poisson noise and temporal jitter inherent in SPAD measuremens [73]: Even though the discontinuity shapes deviate from the ideal shapes in the simulated transient of Figure 3.2, the theoretically predicted features are still visible.

3.3 NLOS Shape Reconstruction Using Fermat Paths

Using the results of Section 3.2, given a transient measurement $I(\tau; v)$, we can identify its discontinuities as the lengths $\tau_{\mathcal{F}}$ of Fermat paths contributing to the transient. Each length $\tau_{\mathcal{F}}$ constrains the corresponding point $x_{\mathcal{F}} \in \mathcal{F}(v)$ to lie on the tangent sphere $\text{Sph}(\tau_{\mathcal{F}}/2; v)$ and, if $x_{\mathcal{F}} \in \mathcal{S}(v)$, also constrains its normal and curvature. Then, given a collection of Fermat pathlengths, we apply the Fermat flow algorithm (Section 2.4.1)



(a) scanning (b) discontinuity detection (c) gradient estimation by interpolation (d) point cloud reconstruction

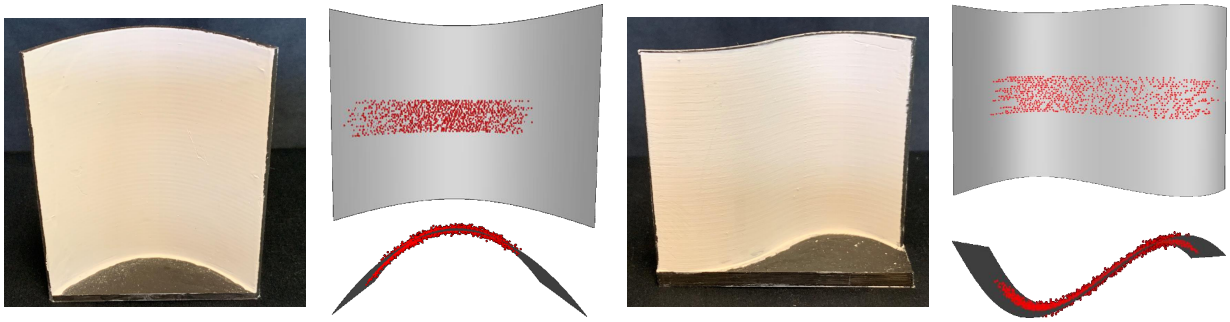
Figure 3.4: **Reconstruction pipeline.** (a) We first collect transient measurements $I(\tau; v)$ at multiple points v on the visible surface $\mathcal{V}$. (b) For each measured transient, we detect pathlengths where the transient is discontinuous. These correspond to samples of the multi-valued Fermat pathlength function $\tau_{\mathcal{F}}(v)$. In the example shown, $\tau_{\mathcal{F}}(v)$ has two branches $\tau_{\mathcal{F},1}(v)$ and $\tau_{\mathcal{F},2}(v)$, shown in blue and red respectively. (c) Within each branch of $\tau_{\mathcal{F}}(v)$, we interpolate to compute the gradient $\nabla_v \tau_{\mathcal{F}}(v)$. (d) Finally, by applying the Fermat flow equation (2.23), we reconstruct from each branch a set of points, either on the boundary (branch $\tau_{\mathcal{F},1}(v)$, blue) or at the interior (branch $\tau_{\mathcal{F},2}(v)$, red) of the NLOS shape.

to produce an *oriented* point cloud (locations and normals) for the NLOS surface $\mathcal{X}$, and then perform surface fitting to obtain a surface reconstruction. Figure 3.4 visualizes our reconstruction pipeline.

Algorithmic details. We detect discontinuities in each measured transient $I(\tau; v)$ with a basic one-dimensional edge detection procedure, by filtering each transient with a set of derivative-of-Gaussian filters, and performing non-max-suppression. To compute gradients $\nabla_v \tau_{\mathcal{F}}(v)$ in the Fermat flow algorithm, we use quadratic interpolation on Fermat pathlength values at a 5×5 neighborhood around each point v.

3.4 Experiments

We discuss results from NLOS reconstruction experiments we have performed to validate and evaluate the Fermat flow algorithm. All experiments are based on measurements captured with two transient imaging setups, one operating at picosecond and the other at femtosecond temporal scales.



(a) photograph and reconstruction of paraboloid object (b) photograph and reconstruction of sigmoid object

Figure 3.5: **Comparison with groundtruth.** We perform one-dimensional scans of 3D-printed objects, in a looking-around-the-corner configuration. For each object, we show a photograph under ambient light (left), and reconstruction results (red points) superimposed against the groundtruth mesh used to fabricate the object (middle and right).

3.4.1 Picosecond-scale Experiments

Imaging system. We use a SPAD-based transient imaging system [55, 133, 135], consisting of a picosecond laser (NKT SuperK EXW-12), a SPAD detector (MPD module), and a time-correlated single photon counter (TCSPC, PicoQuant PicoHarp). The temporal binning resolution of the TCSPC unit is 4 ps, for an absolute upper bound in depth resolution of 1.2 mm. In practice, the resolution is lower, because of laser and TCSPC jitter. We use galvo mirrors to independently control viewpoint and illumination direction, and perform both confocal and non-confocal scanning in the looking-around-the-corner setting of Figure 3.1(a).

Comparison with ground truth. We fabricated small objects from CAD meshes, providing us with ground-truth shape for comparison. The objects were painted with matte white paint to create Lambertian reflectance. The objects are approximately 15 cm in each dimension, and are placed at a distance 25 cm from a planar visible surface. All objects are ruled surfaces, to allow reconstruction of their profile from only one-dimensional scans. We capture measurements under a non-confocal setting, by fixing the point imaged by the SPAD and scanning the point illuminated by the source along multiple horizontal lines on the visible wall. Along each line, we scan 200 points, at a distance of approximately 1 mm from each other. Figure 3.5 shows point clouds reconstructed from these measurements using the Fermat flow procedure, superimposed against the meshes used for fabrication. The reconstructions closely reproduce

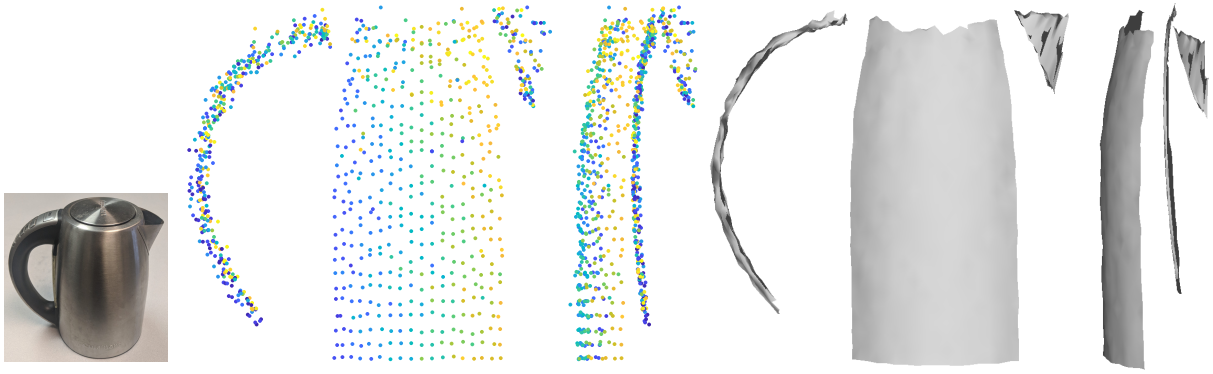
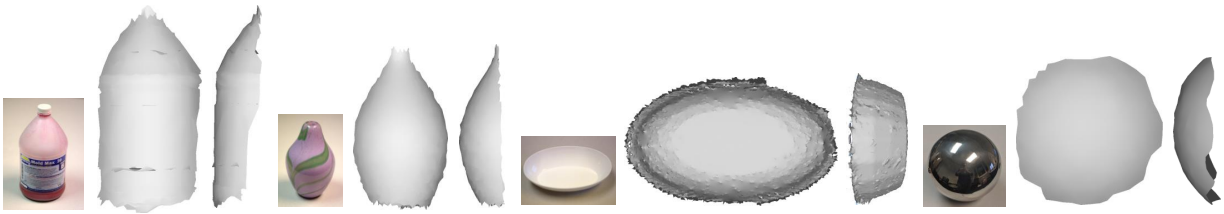


Figure 3.6: **Comparison of point cloud and surface reconstructions.** We scan a rough specular kettle, shown in the left under ambient light. We reconstruct an oriented point cloud, shown in the middle from two views, where points are colored according to their normal. Finally, we fit a surface to the point cloud, shown to the right under two views.



(a) plastic jug (b) glass vase (c) plastic bowl (d) metal sphere

Figure 3.7: **Table-top objects.** We scan objects that span a variety of shapes (convex, concave) and reflectances (translucent, glossy, specular). For each object, we show a photograph under ambient light, and two views of its surface reconstruction.

the shape of the objects, including their concave and convex surfaces, and match the groundtruth within 2 mm.

Table-top objects. We scanned a variety of every day objects (Figures 3.6 and 3.7), with convex and concave geometry of different BRDFs, including translucent (plastic jug), glossy (bowl, vase), rough specular (kettle) and smooth specular (sphere). Most of the objects have a major dimension of approximately $20 - 30$ cm, and are placed at a distance of 80 cm from the visible wall. We use confocal scanning with a grid of 64×64 points distributed in an area of 80 cm $\times$ 80 cm on the visible wall.

Figure 3.6 visualizes point cloud, normal, and final surface reconstruction for one of the objects, an electrical kettle with rough-specular reflectance. We observe that our reconstruction procedure produces a point cloud that closely matches the shape of the ob-

ject, including accurate normals on its front surface. We note that we do not reconstruct normals at the handle of the object: This is expected, because these parts of the object produce Fermat paths of boundary, rather than specular, type, and such paths do not provide normal information. The final fitted surface further improves the reconstruction quality. Figure 3.7 shows reconstructions for the remaining table-top objects. In all cases, the reconstruction closely matches the object shape, demonstrating the ability of our algorithm to handle a variety of complex geometry and reflectance combinations.

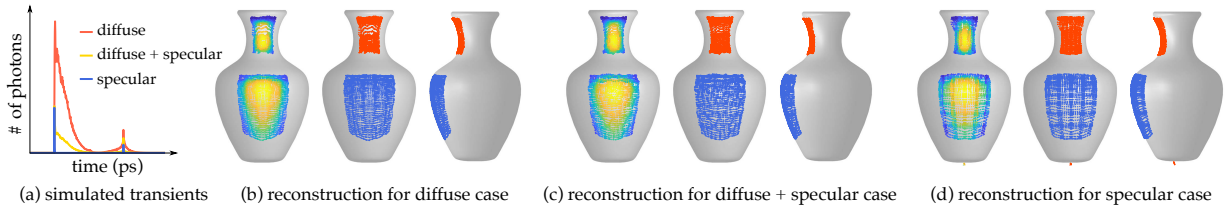
3.4.2 Femtosecond-scale Experiments

Imaging system. We use a time-domain, full-frame optical coherent tomography system [56]. We use this system to perform confocal scans under both the looking-around-the-corner and looking-through-diffuser settings (Figure 3.1). We use spatially and temporally incoherent LED illumination, which allows us to combine transient imaging with diagonal probing [132]. In the context of confocal scanning, this means that we can simultaneously collect transients $I(\tau, v)$ at all points on the visible surface without scanning, as transient measurements taken at one point will not be contaminated with light emanating from a different point. Our implementation has depth resolution of 10μm.

Coin reconstructions. We perform experiments in both looking-around-the-corner and looking-through-diffuser settings (Figure 3.1), where for the diffuser we use a thin sheet of paper. In both cases, the NLOS object is a US quarter, with the obverse side facing the visible surface. We place the coin at a distance of 10 mm from the visible surface, and collect transient measurements on an area of about 40 mm $\times$ 40 mm, at an 1 MPixel grid of points. For validation, we additionally use the same setup to directly scan the coin without occlusion. Figure 3.1 shows our results. In both cases, we can reconstruct fine detail on the coin, sufficient to infer its denomination. The reconstructed detail is also in close agreement with the groundtruth shape measured with the coin directly in the line-of-sight.

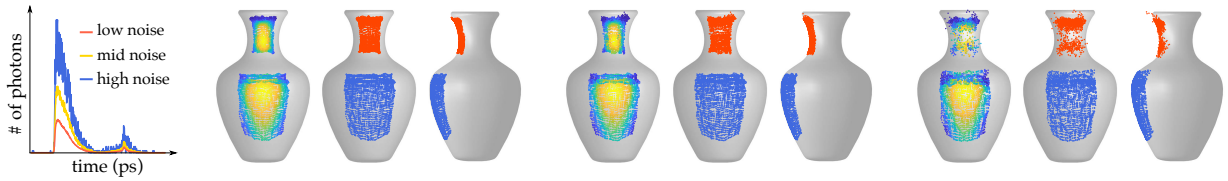
3.4.3 Simulated Experiments

We also performed simulated experiments to evaluate the performance of our reconstruction algorithm under different BRDF and noise conditions. In all cases, the syn-



(a) simulated transients (b) reconstruction for diffuse case (c) reconstruction for diffuse + specular case (d) reconstruction for specular case

Figure 3.8: **Reconstructions under different BRDFs.** We show reconstructions from simulated transient measurements for a vase, rendered under three different BRDFs: Lambertian, mixture of Lambertian and specular, and specular. (a) Comparison of rendered transients for the three cases (representative sample). (b)-(d) For each case, we show reconstructed points colored by normal (left) and by branch of the Fermat pathlength function (middle and right).



(a) simulated transients (b) reconstruction for low noise case (c) reconstruction for mid noise case (d) reconstruction for high noise case

Figure 3.9: **Reconstructions under different noise levels.** We show reconstructions from simulated transient measurements for a vase, rendered with three different noise levels. (a) Comparison of rendered transients for the three cases (representative sample). (b)-(d) For each case, we show reconstructed points colored by normal (left) and by branch of the Fermat pathlength function (middle and right).

thetic transient data was simulated using physically-accurate Monte Carlo rendering [80], and noise was added using the SPAD model of Hernandez *et al*. [73].

Reconstructions under different BRDFs and noise levels. Figures 3.8 and 3.9 show simulated experiments for a vase object. We render 64×64 transients under a confocal scanning scheme. The Fermat pathlength function for the vase object has two branches, both specular, one corresponding to the convex body of the vase, and another to the concave neck of the vase.

In Figure 3.8, we compare rendered data and reconstructions for three different BRDFs, ranging from fully Lambertian to fully specular. We observe that both the reconstructed points and normals remain largely invariant to the BRDF change, as expected from our theory. In Figure 3.9, we compare rendered data and reconstructions for three different noise levels. We observe again that the reconstructions of both points and nor-

mals remain robust as the noise increases.

3.5 Discussion and Summary

We discuss some limitations of our approach. Our reconstruction procedure based on the theory of Fermat paths does not require radiometric calibration, as it does not use intensity information, instead relying on estimating the pathlengths where measured transients are discontinuous. Consequently, our reconstructions can be sensitive to inaccurate discontinuity detection. Reconstruction quality can additionally suffer if we do not have sufficiently *dense* measurements for estimating Fermat pathlength gradients through interpolation, i.e., when *differential* imaging is not adopted. Finally, using only pathlength information provides BRDF invariance, but it also means that we do not take advantage of the information available in measured intensities about the NLOS scene.

The theory of Fermat paths offers new insights into the NLOS imaging problem, linking it to classical areas such as specular and differential geometry, and providing ample opportunity for transfer of ideas from these areas to the NLOS imaging setting. By allowing us to treat NLOS reconstruction from a purely geometric perspective, our theory introduces a new methodology for tackling this problem, distinct from but complementary to approaches such as (elliptic) backprojection [135, 179] and analysis-by-synthesis [176], which focus on the radiometric aspects of the problem.

Interestingly, concurrent work [112] has uncovered an intriguing link between our and backprojection approaches, by showing that the latter cannot reconstruct NLOS points not on Fermat paths, even if those points otherwise contribute to measured transients. Therefore, both approaches reproduce the same part of the NLOS scene. Further exploration of connections between the geometric and backprojection approaches can help shed light into their fundamental limits and strengths, potentially by allowing us to derive results applicable to both classes of approaches using whichever mathematical framework (geometric, radiometric) is more convenient for analysis.

More broadly, an exciting future direction of research is combining the two classes of approaches, not only for NLOS imaging, but also for other related applications, including acoustic and ultrasound imaging [108], lensless imaging [188], and seismic imaging [161].

Chapter 4

Fermat Paths for Specular Shape Reconstruction

In Chapter 3, we have discussed how we apply the theory of Fermat paths for NLOS shape reconstruction. In this chapter, we apply the theory to specular (mirror-like) object reconstruction, this time in a line-of-sight setting. Because the law of specular reflection is derived from Fermat's principle, all light paths reflecting off a specular object are Fermat paths (more precisely, specular paths, and not boundary ones). This allows us to directly measure information of Fermat paths and apply the reconstruction algorithms in Chapter 2.

In Section 4.1, we introduce the imaging setup. In Section 4.2, we explore the geometric information available in the specularities captured by a differentially moving camera. Zisserman *et al*. [198] and our derived integral form of Fermat flow algorithm have shown that the locations of moving specularities can be used for reconstructing a one-parameter family (OPF) of specular surfaces. The algorithm is exactly the integral form of the Fermat flow algorithm. The remaining ambiguity influences the reconstructed shape significantly to be of use to computer vision applications. In order to determine a unique solution from the OPF, regularization is required, e.g., by exploiting prior knowledge of the mirror shape or a known point on the true surface. However, it is often the case that the surface shape is not known in advance and explicitly determining one point on the surface requires extra efforts, e.g., by coating the surface with diffuse paint and use methods for diffuse object reconstruction.

Inspired by shape from shading methods that target surface reconstruction in the presence of specular highlights, e.g., surfaces with both diffuse and specular compo-

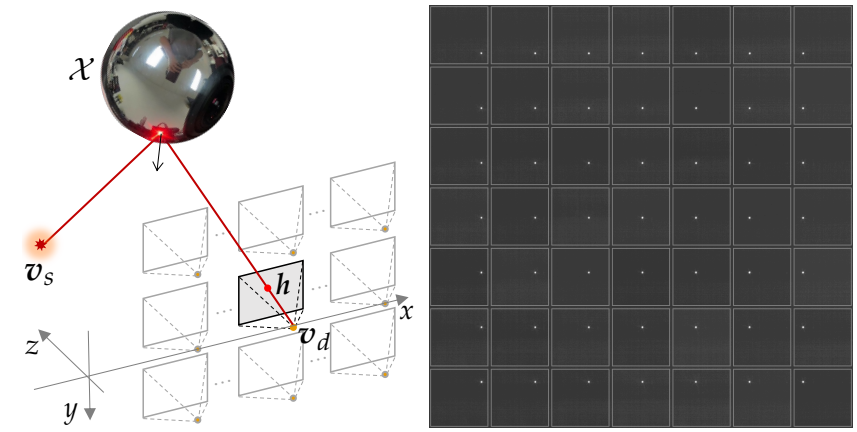
nents or rough specular surfaces [13, 14, 79, 119, 142, 146], we address the following problem: What does radiometric information of specularities observed by a moving camera tell us about the shape of a *purely specular mirror* and can it help reduce the ambiguity within the aforementioned OPF?

In Section 4.3, we explore the radiometric information of moving specularities for resolve the ambiguity in the OPF. We conduct theoretical analysis and prove that a pair of image irradiances of specularities observed at two camera locations can be used to reduce the ambiguity down to a maximum of 16 discrete possibilities for general shapes. By considering all available irradiances, we develop an optimization approach to empirically eliminate the ambiguity to determine the true surface. Our analysis does not require a careful radiometric calibration and is only based on the assumption that the mirror has a homogeneous specular albedo and the incident angles for all measured points on the mirror do not vary much such that the Fresnel term in all reflection events are assumed to be constant.

In Section 4.4, We perform simulations to support our theory on planar and curved mirrors. We verify the correctness of the specularity irradiance model by comparing it against images rendered with a Monte Carlo ray tracer [80]. We then demonstrate that using our proposed optimization approach, radiometric information of moving specularities can be used as a regularization to empirically determine a unique surface within the OPF.

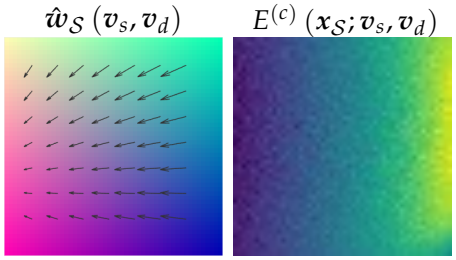
4.1 Imaging Setup

Our imaging setup, shown in Figure 4.1(a), comprises a near-field point light source and a camera, located at $\boldsymbol{v}_s, \boldsymbol{v}_d \in \mathbb{R}^3$, respectively. The camera $\boldsymbol{v}_d$ is differentially translated along the two directions orthogonal to its optical axis and sampled on a dense 2D rectangular grid. The moving camera essentially emulates a multi-camera array/light field camera/wavefront sensor. This setting corresponds to the non-confocal case discussed in Section 2.5. We assume the scene $\mathcal{X} \subset \mathbb{R}^3$ is a piecewise smooth mirror of general unknown shapes placed in close proximity to the imaging setup comparable to its scale. We also assume that light emitted from the point source bounces only once on the specular surface before reaching the sensor and ignore higher-order bounces, such as interreflections. At each camera location $\boldsymbol{v}_d$, the captured image will exhibit one or

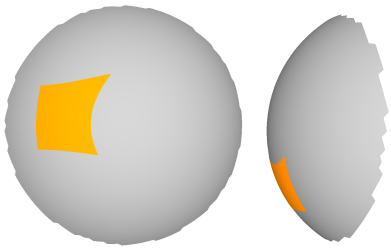


(a) imaging setup

(b) specularities



(c) geometric and radiometric measurements



(d) surface reconstruction

Figure 4.1: **Near-light differential imaging for specular object reconstruction.** (a) Our imaging setup consists of a near-field point light source v_s and a translating camera v_d, which emulates a multi-camera array with optical centers sampled on a 2D regular grid. We assume the object has a smooth mirror reflectance. (b) At each camera location v_d, the captured image will exhibit one or more specular highlights, corresponding to some specular paths $v_s \to x_S \to v_d$. In this specific example, there is only one highlight h within each image. (c) Zisserman *et al*. [198] examines the geometric information associated with these specularities, i.e., the directions of specular reflections $\hat{w}_S$, and show that they can be integrated to obtain a one-parameter family (OPF) of surfaces. We review their work in Section 4.2. In this work, we show that radiometric information, i.e., the absolute image irradiances of specularities, can be used to further disambiguate within the OPF. This is detailed in Section 4.3. (d) We apply surface fitting to the oriented point cloud and obtain the final surface reconstruction. Ground truth surface is shown in gray.

more specularities, corresponding to different specularly reflected rays. By considering images at all camera locations, we obtain a mosaic of specularities (Figure 4.1(b)), whose

pixel locations and image irradiances provide *geometric* and *radiometric* information of the specularly reflected rays (Figure 4.1(c)), respectively. In the example shown in Figure 4.1, there is only one specularity per image. Depending on the specular surface, there could exists more specularities per view. In this case, as discussed in Chapter 2, each branch of specularities should be processed separately.

4.2 Geometric Information from Specularities

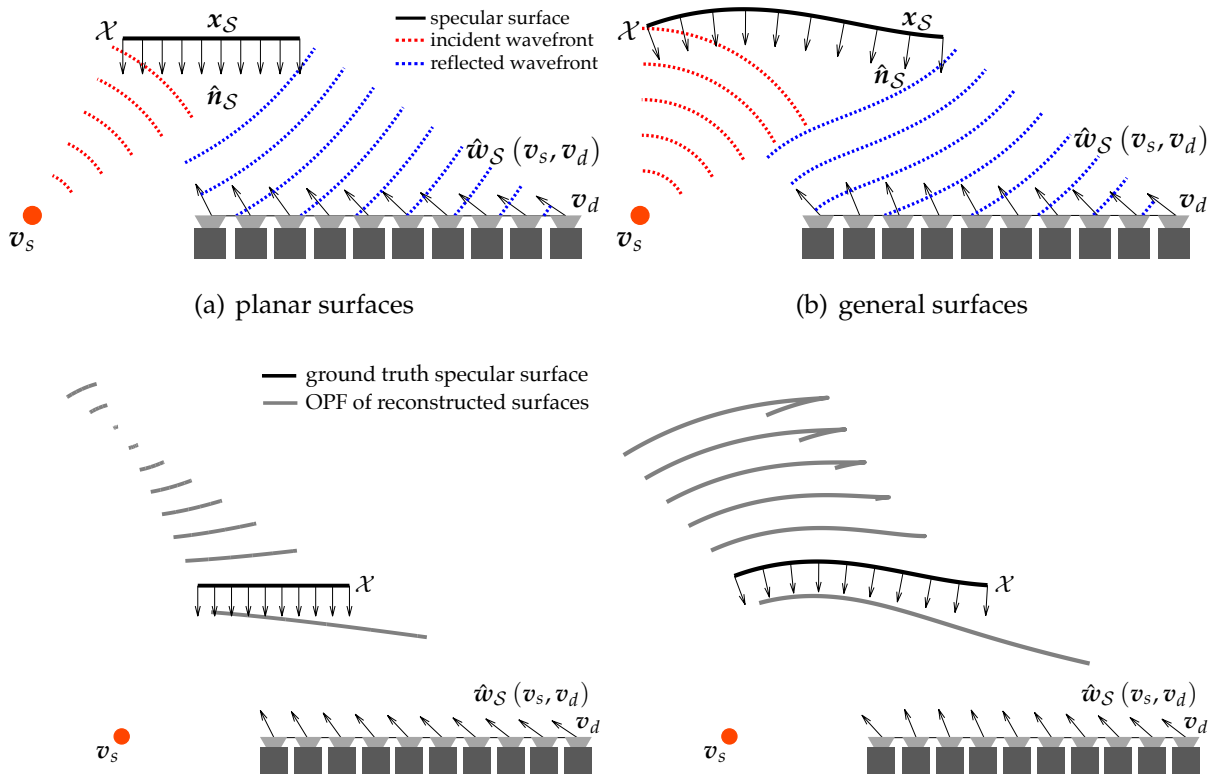
We first explore how geometric information of moving specularities help reconstruct the specular surface. Let path $\boldsymbol{p}(\boldsymbol{x}; \boldsymbol{v}_s, \boldsymbol{v}_d)$ be a specular path. Assume the reflected ray $\boldsymbol{x}_{\mathcal{S}} \rightarrow \boldsymbol{v}_d$ creates a highlight on the image plane with homogeneous coordinate $\boldsymbol{h} = M(\boldsymbol{x}_{\mathcal{S}} - \boldsymbol{v}_d)$, where M denotes the camera's intrinsic matrix. Given M, we can back-project $\boldsymbol{h}$ to the specularly reflected ray: $\boldsymbol{w}_{\mathcal{S}}(\boldsymbol{v}_s, \boldsymbol{v}_d) = \boldsymbol{x}_{\mathcal{S}} - \boldsymbol{v}_d = M^{-1}\boldsymbol{h}$. Further normalizing $\boldsymbol{w}_{\mathcal{S}}$ yields the unit direction $\hat{\boldsymbol{w}}_{\mathcal{S}}(\boldsymbol{v}_s, \boldsymbol{v}_d) = \frac{\boldsymbol{w}_{\mathcal{S}}(\boldsymbol{v}_s, \boldsymbol{v}_d)}{\|\boldsymbol{w}_{\mathcal{S}}(\boldsymbol{v}_s, \boldsymbol{v}_d)\|}$. $\hat{\boldsymbol{w}}_{\mathcal{S}}(\boldsymbol{v}_s, \boldsymbol{v}_d)$ is the surface normal of the reflected wavefront (shown as blue dotted curves in Figure 4.2(a) and (b)), or equivalently the gradient of the FPF (Equation (2.43)).

Given one branch of directional measurements of specularly reflected rays $\hat{\boldsymbol{w}}_{\mathcal{S}}(\boldsymbol{v}_s, \boldsymbol{v}_d)$, we can apply the integral form of the Fermat flow algorithm (Proposition 9′ in Chapter 2) to reconstruct a OPF of specular surfaces. We restate the equation here for convenience:

$$\tau_{\mathcal{S}}(\boldsymbol{v}_s, \boldsymbol{v}_d; \tau_0) = \int_{\boldsymbol{v}_d} -\hat{\boldsymbol{w}}_{\mathcal{S}}(\boldsymbol{v}_s, \boldsymbol{v}_d)\, \mathrm{d}\boldsymbol{v}_d + \tau_0 \,. \tag{4.1}$$

Note that we explicitly include the unknown constant τ_0 in the parameterization. Each τ_0 determines one reflected wavefront, which can be converted to the specular surface $\mathcal{X}$ through an ellipsoid-ray intersection (Section 2.5.3). In the context of reconstructing specular surfaces from moving specularities, Equation (4.1) is closely realted to the theorem proposed by Zisserman *et al*. [198].

The influence of the unknown pathlength offset τ_0 on specular shape reconstruction. Without resolving τ_0, the OPF of reflected wavefronts in our near-light setup gets converted to a OPF of specular surfaces that of dramatically different shapes (Figure 4.2(c) and (d)). Specifically, for a planar surface, an incorrect τ_0 leads to a surface reconstruction that is no longer planar, but rather be convex or concave. The surface can even degenerate and collapse into a point (which is exactly the mirror reflection of the light source $\boldsymbol{v}_s$ about the plane). For general shapes, the reconstructed OPF contains surfaces



(a) planar surfaces

(b) general surfaces

(c) OPF of reconstructed specular surfaces

(d) OPF of reconstructed specular surfaces

Figure 4.2: **Geometric wavefronts and one-parameter family (OPF) of specular surface reconstruction**. A point light source v_s emits a spherical wavefront (dotted curves in red) towards the specular surface. After specular reflection, the reflected wavefront (dotted curves in blue) gets modulated by the specular surface. (a) For planar surfaces, the reflected wavefront does not change its shape and remains spherical. (b) For surfaces of general shapes, the reflected wavefront evolves into a new shape. In both cases, the reflected wavefront propagates to the translating camera (shown as a multi-camera array). The measured directions of specular reflections $\hat{w}_{\mathcal{S}}(v_s, v_d)$ can be integrated to reconstruct a OPF of reflected wavefronts (Equation (4.1)), which can then be converted to a OPF of specular surfaces. The OPF of specular surfaces do not share the same shape, as shown in (c) and (d), which motivates additional regularizations to uniquely determine the surface.

that can be discontinuous and "grow" new patches. We can think of it as the inverse of converging rays forming caustics; this time, a point can expand into a new surface patch as τ_0 changes.

Because of the dramatically varying shapes in the OPF, it is important to uniquely determine the surface for the reconstruction to be of practical use, e.g., controlling the quality of manufactured mirror objects. To resolve the ambiguity, additional regularization is needed, which remains an open-research problem [28]. Zisserman *et al*. [198] propose to further locate one point on the specular surface. However, explicitly determining one point on a mirror requires extra efforts, e.g., by coating the object with diffuse paint and applying reconstruction methods for Lambertian objects.

4.3 Radiometric Information from Specularities

In this section, we show it is possible to use radiometric information of specularities to reduce the number of feasible solutions of τ_0 from infinity to a maximum of 16, and empirically down to a unique solution. In Section 4.3.1, we derive image irradiances of specularities involving a single reflection event and show that they encode curvature information of the specular surfaces. Through a reparameterization described in Section 4.3.2, we express image irradiances as a function of the pathlenght offset τ_0. In Section 4.3.3, we perform a theoretical analysis on the uniqueness of solutions when using radiometric information for resolving τ_0.

4.3.1 Image Irradiances of Specularities

Our derivation follows prior works [122, 168] for tracing multiple specular reflection or refraction events. We adapt the derivation to our setup involving only a single reflection from a point source.

Proposition 12. *Let us consider the specular path* $\boldsymbol{v}_s \rightarrow \boldsymbol{x}_{\mathcal{S}} \in \mathcal{X} \rightarrow \boldsymbol{v}_d$*, and denote the Gaussian curvature of the wavefront: (1) emitted from the point light source to* $\boldsymbol{x}_{\mathcal{S}}$ *as* $K^{(i)}$*, (2) immediately after reflection as* $K^{(r)}$*, and (3) further propagating to the camera as* $K^{(c)}$*, where the superscripts* (i)*,* (r) *and* (c) *denote incident, reflected, and wavefronts arriving at the camera, respectively.*

Let $I_{\boldsymbol{v}_s}$ *be the radiant intensity of the near-field point light source. Let* θ_i *and* θ_c *be the incident angle of specular reflection, and the angle between the specularly reflected ray and camera viewing direction, respectively. Let* ρ *be the specular albedo of the smooth homogeneous specular surface. Let* $F_r(\theta_i)$ *be the Fresnel term dependent on the incident angle. Then, the image*

irradiance of the specular highlight can be expressed as

$$E^{(c)} = F_r\left(\theta_i\right) \rho I_{\boldsymbol{v}_s} K^{(i)} \frac{\left|K^{(c)}\right|}{\left|K^{(r)}\right|} \cos\left(\theta_c\right) . \tag{4.2}$$

Proof. The key idea is to keep track of the Gaussian curvature change of the wavefront as it propagates because irradiance at any point is proportional to the Gaussian curvature of the wavefront at that point [122]. The tracking consists of three steps. First, the point light source $\boldsymbol{v}_s$ emits an initial spherical wavefront and propagates to the specular surface at point $\boldsymbol{x}_{\mathcal{S}}$. Second, immediately after reflection, the wavefront gets modulated by the specular surface. Third, the reflected wavefront further propagates to the camera $\boldsymbol{v}_d$.

Incident wavefront. Given the radiant intensity of the near-field point light source $I_{\boldsymbol{v}_s}$, we compute the irradiance of the initial spherical wavefront at point $\boldsymbol{x}_{\mathcal{S}}$ according to the inverse-square law, also known as the quadratic falloff:

$$E^{(i)} = \frac{I_{\boldsymbol{v}_s}}{\left\|\boldsymbol{x}_{\mathcal{S}} - \boldsymbol{v}_s\right\|^2} = I_{\boldsymbol{v}_s} K^{(i)} , \tag{4.3}$$

where $K^{(i)} = 1/\left\|x_{\mathcal{S}} - v_s\right\|^2$ is the Gaussian curvature of the incident spherical wavefront arriving at $\boldsymbol{x}_{\mathcal{S}}$.

Reflected wavefront. After reflection, the wavefront evolves into a new shape which is dependent on the shape of the specular surface. According to the law of conservation of energy, an idealized 100% reflective mirror reflects all received flux in the outgoing direction. In practice, we need to also consider specular albedo ρ and Fresnel coefficient $F_r\left(\theta_i\right)$, which leads us to the following:

$$E^{(r)} = F_r\left(\theta_i\right) \rho E^{(i)} . \tag{4.4}$$

Reflected wavefront propagating to the camera. Finally, we propagate the reflected wavefront in the free space towards the camera. Again, based on conservation of energy, the irradiance along reflected ray is proportional to the Gaussian curvature, therefore

$$\frac{E_{\perp}^{(c)}}{E^{(r)}} = \frac{\left|K^{(c)}\right|}{\left|K^{(r)}\right|} , \tag{4.5}$$

where $E_{\perp}^{(c)}$ denotes the wavefront irradiance along the reflected ray when arriving at the camera. The $\perp$ subscript represents the irradiance received by a sensor placed perpendicularly to the outgoing ray direction. A derivation of Equation (4.5) can be found in Mitchell and Hanrahan [122].

When the camera viewing direction and the reflected ray direction span an angle θ_c, we account for the foreshortening term $\cos(\theta_c)$:

$$E^{(c)} = E_{\perp}^{(c)} \cos(\theta_c) \,. \tag{4.6}$$

Combining Equations (4.3)-(4.6) completes the proof. □

4.3.2 Reparameterization of Equation (4.2)

We reparameterize Equation (4.2) using $\boldsymbol{v}_s, \boldsymbol{v}_d$ and the unknown pathlength offset τ_0. First, simplifying the notation leaving the curvature terms yields

$$\tilde{E}^{(c)} = \tilde{I}_{\boldsymbol{v}_s} K^{(i)} \frac{\left|K^{(c)}\right|}{\left|K^{(r)}\right|} \,, \tag{4.7}$$

where $\tilde{E}^{(c)} = E^{(c)}/\cos(\theta_c)$ denotes image irradiances absorbing the camera orientation and $\tilde{I}_{\boldsymbol{v}_s} = F_r(\theta_i)\,\rho I_{\boldsymbol{v}_s}$ is the scaled light source intensity absorbing material properties.

The curvature terms $K^{(i)}$ and $\left|K^{(c)}\right|/\left|K^{(r)}\right|$ can be expressed using four geometric quantities, including the incident and outgoing pathlengths denoted as $\tau_{\mathcal{S}}^{(i)}$ and $\tau_{\mathcal{S}}^{(r)}$, as well as the Gaussian and mean curvatures of the reflected wavefront arriving at the camera, denoted as $K^{(c)}$ and $H^{(c)}$. All four geometric quantities can be computed by integrting or differentiating $\hat{\boldsymbol{w}}_{\mathcal{S}}(\boldsymbol{v}_s, \boldsymbol{v}_d)$, whose derivations were already provided in Equations (2.47), (2.48), (2.32), and (2.34), respectively. We restate their expressions as follows. Note that we explicitly include τ_0 in the parameterization.

$$\tau_{\mathcal{S}}^{(i)}(\boldsymbol{v}_s, \boldsymbol{v}_d; \tau_0) = \frac{{\tau_{\mathcal{S}}}^2 + 2\langle \boldsymbol{v}_d - \boldsymbol{v}_s, \hat{\boldsymbol{w}}_{\mathcal{S}} \rangle \tau_{\mathcal{S}} + \|\boldsymbol{v}_d - \boldsymbol{v}_s\|^2}{2\left(\tau_{\mathcal{S}} + \langle \boldsymbol{v}_d - \boldsymbol{v}_s, \hat{\boldsymbol{w}}_{\mathcal{S}} \rangle\right)} \,, \text{ and} \tag{4.8}$$

$$\tau_{\mathcal{S}}^{(r)}(\boldsymbol{v}_s, \boldsymbol{v}_d; \tau_0) = \frac{{\tau_{\mathcal{S}}}^2 - \|\boldsymbol{v}_d - \boldsymbol{v}_s\|^2}{2\left(\tau_{\mathcal{S}} + \langle \boldsymbol{v}_d - \boldsymbol{v}_s, \hat{\boldsymbol{w}}_{\mathcal{S}} \rangle\right)} \,. \tag{4.9}$$

$$K^{(r)}(\boldsymbol{v}_s, \boldsymbol{v}_d; \tau_0) = \frac{K^{(c)}}{1 - 2H^{(c)}\tau_{\mathcal{S}}^{(r)} + K^{(c)}{\tau_{\mathcal{S}}^{(r)}}^2} \,. \tag{4.10}$$

Plugging Equation (4.10) and $K^{(i)} = 1/{\tau_{\mathcal{S}}^{(i)}}^2$ to Equation (4.7), we obtain a reparameterization of irradiance expressed in terms of the unknown pathlength offset τ_0:

$$\tilde{E}^{(c)}(\boldsymbol{v}_s, \boldsymbol{v}_d; \tau_0) = \tilde{I}_s \frac{\left|1 - 2H^{(c)}\tau_{\mathcal{S}}^{(r)} + K^{(c)}{\tau_{\mathcal{S}}^{(r)}}^2\right|}{{\tau_{\mathcal{S}}^{(i)}}^2} \,. \tag{4.11}$$

4.3.3 Resolving τ_0 using Radiometric Information of Specularities

Equation (4.11) indicates that irradiances of specularities are dependent on τ_0 by relating geometric quantities of the wavefront. As a result, we propose to invert the process and recover τ_0 given radiometric measurements $\tilde{E}^{(c)}$.

In this work, we assume that the specular object has a homogeneous specular albedo ρ and the angle of incidence θ_i are all within a certain range excluding extreme angles (usually less than 60° for most metal materials) such that the Fresnel term $F_r(\theta_i)$ remains approximately constant. Therefore, the scaled light source intensity $\tilde{I}_{v_s} = F_r(\theta_i)\rho I_{v_s}$ now becomes a *global* constant. [1] Rearranging Equation (4.7), the following expression is assumed to be constant across all measurements:

$$\tilde{I}_s = \frac{\overbrace{\tilde{E}^{(c)}\tau_{\mathcal{S}}^{(i)}(\tau_0)^2}^{\equiv\mathcal{N}}}{\underbrace{\left|1 - 2H^{(c)}\tau_{\mathcal{S}}^{(r)}(\tau_0) + K^{(c)}\tau_{\mathcal{S}}^{(r)}(\tau_0)^2\right|}_{\equiv\mathcal{D}}}, \tag{4.12}$$

where $\mathcal{N}$ and $\mathcal{D}$ denote the numerator and the denominator (inside the absolute sign), respectively. Therefore, for image irradiances of two specularities observed at two camera locations, $\mathcal{N}_1/|\mathcal{D}_1| = \mathcal{N}_2/|\mathcal{D}_2|$. Or equivalently,

$$\mathcal{N}_1\mathcal{D}_2 \pm \mathcal{N}_2\mathcal{D}_1 = 0\,. \tag{4.13}$$

Plugging Equations (4.8) and (4.9) to Equation (4.13) and noting that numerators of both $\tau_{\mathcal{S}}^{(i)}$ and $\tau_{\mathcal{S}}^{(r)}$ are second-order polynomials of τ_0, we have Equation (4.13) as two eighth-order polynomials in terms of τ_0. The roots are potential solutions of τ_0, therefore the ambiguity is reduced down to 16 given a pair of image irradiances.

As will be demonstrated in simulated experiments, empirically we do not get 16 feasible solutions, mainly due to two reasons. First, potential solutions of τ_0 can be imaginary, or real but physically infeasible, e.g., an offset leading to a surface reconstruction behind the camera plane. Second, by taking measurements with a differentially translating camera (or a light field camera/wavefront sensor), we usually have not only two but much more irradiance measurements. For example, a spatial sampling of m provides $(m-1)$ unique irradiance pairs, and we can write a system of $(m-1)$ constraints

[1] When the assumption fails, e.g., when there exist extreme angles of incidence and the Fresnel term varies across measurements, one can perform a radiometric calibration of the scaled light intensities $\tilde{I}_{v_s}$ by taking measurements of a specular object of a known ground truth shape and composited of the same specular material as the mirror of interest.

in the form of Equation (4.13). Common roots of the entire system will usually be much fewer than 16.

In the following, we first prove that for planar specular surfaces, τ_0 has a *unique* analytical solution given the system of polynomials. We then consider surfaces of general shapes which do not have closed-formed solutions, and propose to solve for τ_0 through an optimization instead.

Planar surfaces: a unique solution. As shown in Figure 4.2(a), a planar specular surface does not change the shape of the incident wavefront after reflection—it merely reflects it. As a result, the Gaussian curvatures before and after reflection remain the same, i.e. $K^{(i)} = K^{(r)}$ and Equation (4.7) can be simplified as

$$\tilde{E}^{(c)}\left(\boldsymbol{v}_s, \boldsymbol{v}_d; \tau_0\right) = \tilde{I}_s \left|K^{(c)}\right| = \frac{\tilde{I}_s}{\tau_{\mathcal{S}}^2}\,. \tag{4.14}$$

Equation (4.14) can be derived by following the wavefront tracing procedure described in [122, 168] and noting that the curvature of a plane equals 0. Intuitively, we can imagine placing a virtual point light source at the mirror location of the light source about the planar surface, and directly computing the image irradiance arriving at the camera, which is proportional to the Gaussian curvature of a spherical wavefront $1/{\tau_{\mathcal{S}}}^2$. Equation (4.14) leads us to the following proposition on the uniqueness of surface reconstruction if the ground truth surface is planar. It is important to note that we do not assume the planarity of the mirror, but rather examine whether it is possible to uniquely determine the planar surface within the OPF given radiometric information.

Proposition 13. *Let $\mathcal{X}$ be planar. Let paths $\boldsymbol{v}_s \to \boldsymbol{x}_{\mathcal{S}1} \in \mathcal{X} \to \boldsymbol{v}_{d1}$ and $\boldsymbol{v}_s \to \boldsymbol{x}_{\mathcal{S}2} \in \mathcal{X} \to \boldsymbol{v}_{d2}$ be two specular paths, and $\tau_{\mathcal{S}1}$ and $\tau_{\mathcal{S}2}$ be the ground truth pathlengths, respectively. Assume the pathlengths $\tau_{\mathcal{S}1} + \tau_0$ and $\tau_{\mathcal{S}2} + \tau_0$ also satisfy radiometric constraints (Equation(4.13)). Plugging Equation (4.14) to Equation(4.13) leads to the following seventh-order polynomial in terms of τ_0 :*

$$\tau_0 \left(\tau_0^2 - l^2\right)^2 \left[\tau_0^2 + \left(m_1 + m_2\right)\tau_0 + l^2\right] = 0\,, \tag{4.15}$$

where l is twice the distance from the point light source $\boldsymbol{v}_s$ to $\mathcal{X}$, and m_1 and m_2 are defined as follows:

$$l = 2 \cdot dist\left(\boldsymbol{v}_s, \mathcal{X}\right)\ ,\ and \tag{4.16}$$

$$m_j = \tau_{\mathcal{S}j} + \left\langle \boldsymbol{v}_{dj} - \boldsymbol{v}_s, \hat{\boldsymbol{w}}_{\mathcal{S}}\left(\boldsymbol{v}_s, \boldsymbol{v}_{dj}\right)\right\rangle, j \in \{1, 2\}\ . \tag{4.17}$$

We provide the proof in Appendix C.1. Proposition 13 implies that for a planar surface, τ_0 has five unique solutions: $\left\{0, \pm l, \frac{-(m_1+m_2)\pm\sqrt{(m_1+m_2)^2-4l^2}}{2}\right\}$. The trivial solution $\tau_0 = 0$ yields the ground truth specular shape. $\tau_0 = \pm l$ correspond to zero denominators and are degenerate. The last two solutions are also infeasible because they are imaginary: $m_1, m_2 < l \Rightarrow (m_1 + m_2)^2 - 4l^2 < 0$. As a result, for a planar surface, a pair of radiometric measurements of two specularities already let us resolve τ_0 and uniquely determine the specular surface within the OPF.

General surfaces: an optimization problem. For general surfaces, Equation (4.13) does not have analytical solutions. As a result, we turn to a joint estimation of $\tilde{I}_{v_s}$ and the pathlength offset τ_0. Assuming Gaussian noise in radiometric measurements, we optimize for the two unknowns by minimizing the L2-loss over all specularities:

$$\mathcal{L}\left(\tilde{I}_{v_s}; \tau_0\right) = \left\| \boldsymbol{E}^{(c)} - \tilde{I}_{v_s} \boldsymbol{e}\left(\tau_0\right) \right\|_2 , \tag{4.18}$$

where $e\left(\tau_0\right) = \frac{\left|1-2H^{(c)}\tau_{\mathcal{S}}^{(r)}(\tau_0)+K^{(c)}\tau_{\mathcal{S}}^{(r)}(\tau_0)^2\right|}{\tau_{\mathcal{S}}^{(i)}(\tau_0)^2}\cos\left(\theta_c\right)$, and the bold symbol denotes the vectorization of the respective variables.

The above 2D optimization problem can be transformed into a 1D problem on τ_0 because $\tilde{I}_{v_s}$ is a scalar whose optimal solution is dependent on $\boldsymbol{e}\left(\tau_0\right)$: given any τ_0, minimizing $\mathcal{L}$ over $\tilde{I}_{v_s}$ is a least-square problem with an analytical solution:

$$\tilde{I}_{v_s}^{\star}\left(\tau_0\right) = \frac{\boldsymbol{e}\left(\tau_0\right)^{\top} \boldsymbol{E}^{(c)}}{\boldsymbol{e}\left(\tau_0\right)^{\top} \boldsymbol{e}\left(\tau_0\right)} . \tag{4.19}$$

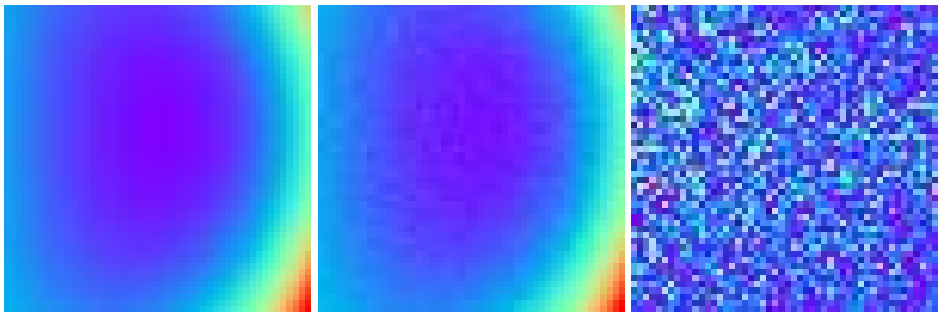
Therefore, minimizing the L2-loss $\mathcal{L}$ is equivalent to maximizing the cosine similarity (minimizing the angle) between $\boldsymbol{e}\left(\tau_0\right)$ and $\boldsymbol{E}^{(c)}$:

$$\min_{\tilde{I}_{v_s}, \tau_0} \mathcal{L}\left(\tilde{I}_{v_s}; \tau_0\right) \tag{4.20}$$

$$\equiv \min_{\tau_0} \mathcal{L}\left(\tilde{I}_{v_s}^{\star}\left(\tau_0\right); \tau_0\right) \tag{4.21}$$

$$\equiv \min_{\tau_0} \mathcal{L}_{\cos}\left(\tau_0\right) = 1 - \left\langle \frac{\boldsymbol{E}^{(c)}}{\left\|\boldsymbol{E}^{(c)}\right\|}, \frac{\boldsymbol{e}\left(\tau_0\right)}{\left\|\boldsymbol{e}\left(\tau_0\right)\right\|} \right\rangle \tag{4.22}$$

The solution to Equation (4.22) is the minimizer of the system of polynomials in the form of Equation (4.13). Because of the non-convexity of the optimization problem, in practice we first conduct a coarse 1D line search over a reasonable range $[\tau_{0\min}, \tau_{0\max}]$ for a good initialization of τ_0, and then refine the initialization with gradient descent.



(a) Equation (4.7) (b) Mitsuba [80] (c) residual

Figure 4.3: **Verifying Equation (4.7) using Monte Carlo rendering.** We verify the correctness of (a) the image irradiance model using (b) images rendered with Mitsuba [80]. (c) The residual (maximum difference less than 6%) shows that our model matches well with Monte Carlo ray tracing.

4.4 Simulated Experiments

In this section, we discuss the simulated experiments we have performed to verify our theory. We start by verifying the image irradiance model (Equation (4.2)) in Section 4.4.1. We then evaluate the uniqueness of τ_0 in Section 4.4.2. Finally, we test the robustness of our optimization method in the presence of radiometric noise in Section 4.4.3. In all simulations, we translate the camera both horizontally and vertically to cover a 51×51 regular grid.

4.4.1 Verifying the Forward Image Irradiance Model

We verify the forward model by comparing it against images rendered with Mitsuba [80], which is a physics-based Monte Carlo ray tracer. We set the reflectance of the object to be slightly rough specular, instead of a smooth mirror reflectance, for rendering stability. Specifically, we choose the GGX microfacet model with roughness 0.001, and select the "none" material profile which produces an idealized 100% reflecting mirror and ignores the computation of the Fresnel term.

Preprocessing. Given rendered images, we detect specular highlights and process them to obtain the geometric and radiometric measurements of the reflected wavefront. Because of the finite image resolution and slightly rough specular reflectance, specularities usually appear as blobs spreading across multiple pixels. For geometric processing, we

compute the centroids of specularities and convert them to directions of specular reflection. For radiometric processing, we compute irradiances of specularities by summing up all image irradiances within the blob. Figure 4.3 shows the image irradinces comparison for a concave sphere. Results show that our forward irradiance model matches well with the Mitsuba rendering, with a residual smaller than 6%.

4.4.2 Uniqueness of τ_0 from Radiometric Information of Two or More Specularities

Next, we explore the uniqueness of τ_0 for general quadrics of the form $z = f_0 + f_x x + f_y y + f_{xx} x^2 + f_{xy} xy + f_{yy} y^2$, which we believe are general enough to represent the *local* shape of most manufactured mirror objects. Specifically, we consider planes, convex and concave paraboloids, and saddles, as shown in Figure 4.4.

We start by exploring the number of feasible roots of Equation (4.13) given *one* pair of randomly picked image irradiances, or equivalently constructing a 2-vector of $\boldsymbol{E}^{(c)}$ and $\boldsymbol{e}(\tau_0)$ in Equation (4.22). We first compute ground truth pathlengths $\tau_S(\boldsymbol{v}_s, \boldsymbol{v}_d)$ for all camera locations, and manually offset them such that the smallest value becomes 0. This step makes sure that negative roots will be physically infeasible because they correspond to surfaces that are behind the camera plane. Consider the blue (or magenta) curve in Figure 4.4(c), which is a polynomial-τ_0 plot. The ground truth offset is shown as the green dot. As we can see, there are multiple zero-crossing points for each polynomial, especially for non-planar surfaces, but many of them are negative and infeasible, which suggests that we have already reduced the ambiguity of τ_0 from infinity to a sparse set of feasible solutions and the size of the feasible set is usually smaller than 16.

We then consider *two* polynomials (both blue and magenta curves) and observe that they rarely share the same set of roots except for the ground truth τ_0. The observation suggests that as we include more radiometric measurements, we can gradually reduce the feasible set of τ_0, which leads us to Figure 4.4(d), where we show the loss curve versus τ_0 when we consider *all* measurements.

These empirical results show that using radiometric measurements we obtain a unique solution of τ_0 for all the shapes we simulated. The zero-crossing points in Figure 4.4(d) also align well with their respective ground truth, suggesting a correct recovery of τ_0, and hence a correct solution of the mirror surface.

4.4.3 Robustness to Gaussian Noise in Radiometric Measurements

To evaluate the robustness of our proposed method in the presence of Gaussian image noise, we add extensive noise (signal-to-noise ratios (SNRs) being 10dB, 7dB, and 5dB, respectively) to the noiseless irradiances computed based on our model, as shown in Figure 4.5. The loss curves plotted in Figure 4.5(e) show that despite the large amount of noise, our method can robustly recover an offset τ_0 that remains very close to the ground truth.

4.5 Discussion and Summary

In this work, we have conducted a theoretical analysis on using radiometric information of specularities as a regularization to the OPF of surfaces obtained using the integral form of the Fermat flow algorithm. We have proven that image irradiances of two specularities already reduce the ambiguity down to a maximum of 16 solutions. We then show through simulated experiments on general quadric shapes that by collectively considering radiometric information of all available specularities, it is possible to empirically get a unique surface reconstruction within the OPF, even in the presence of severe Gaussian noise. In order to bring the method into practical use, we envision as a future work to conduct real experiments to further verify the proposed method.

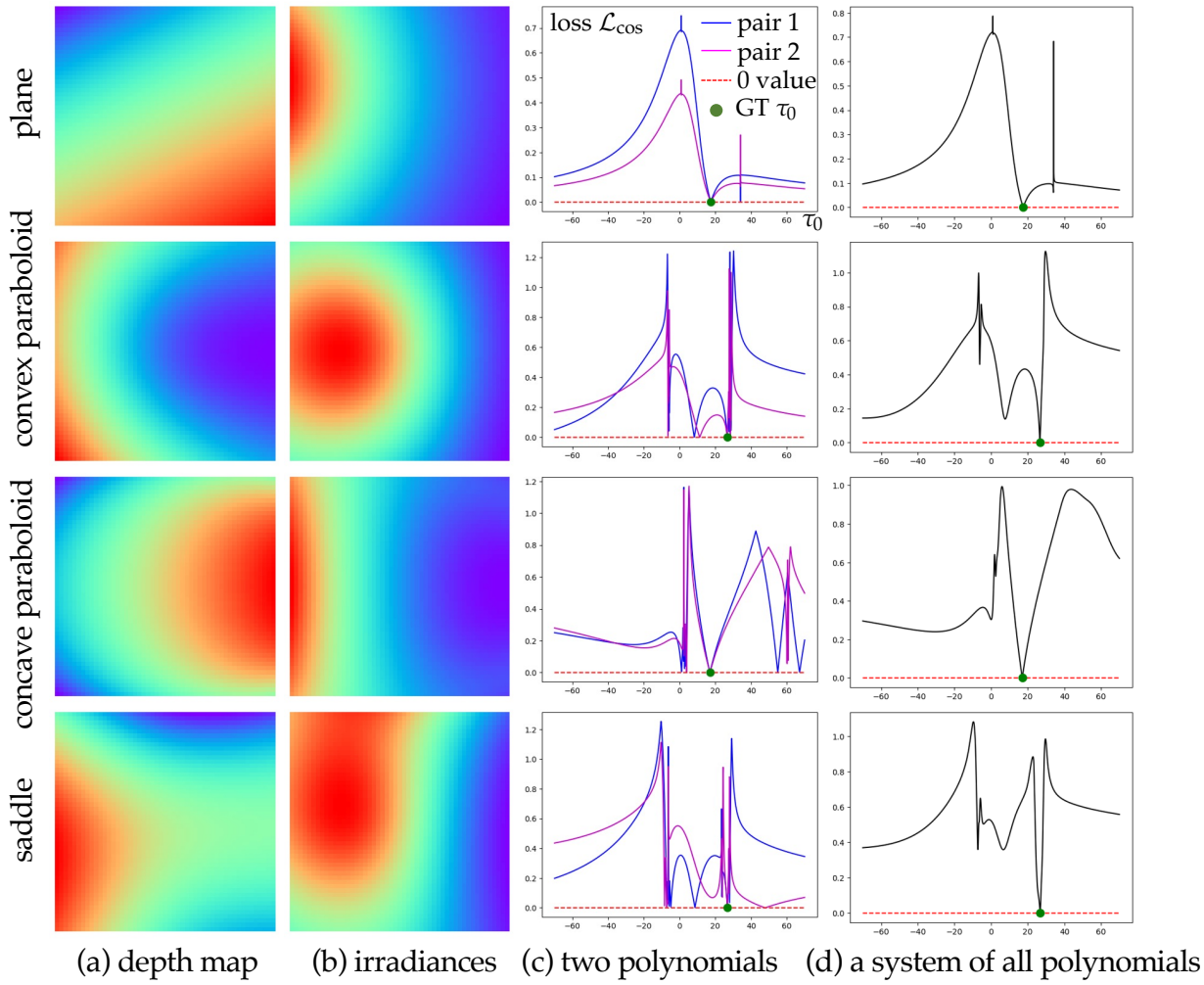
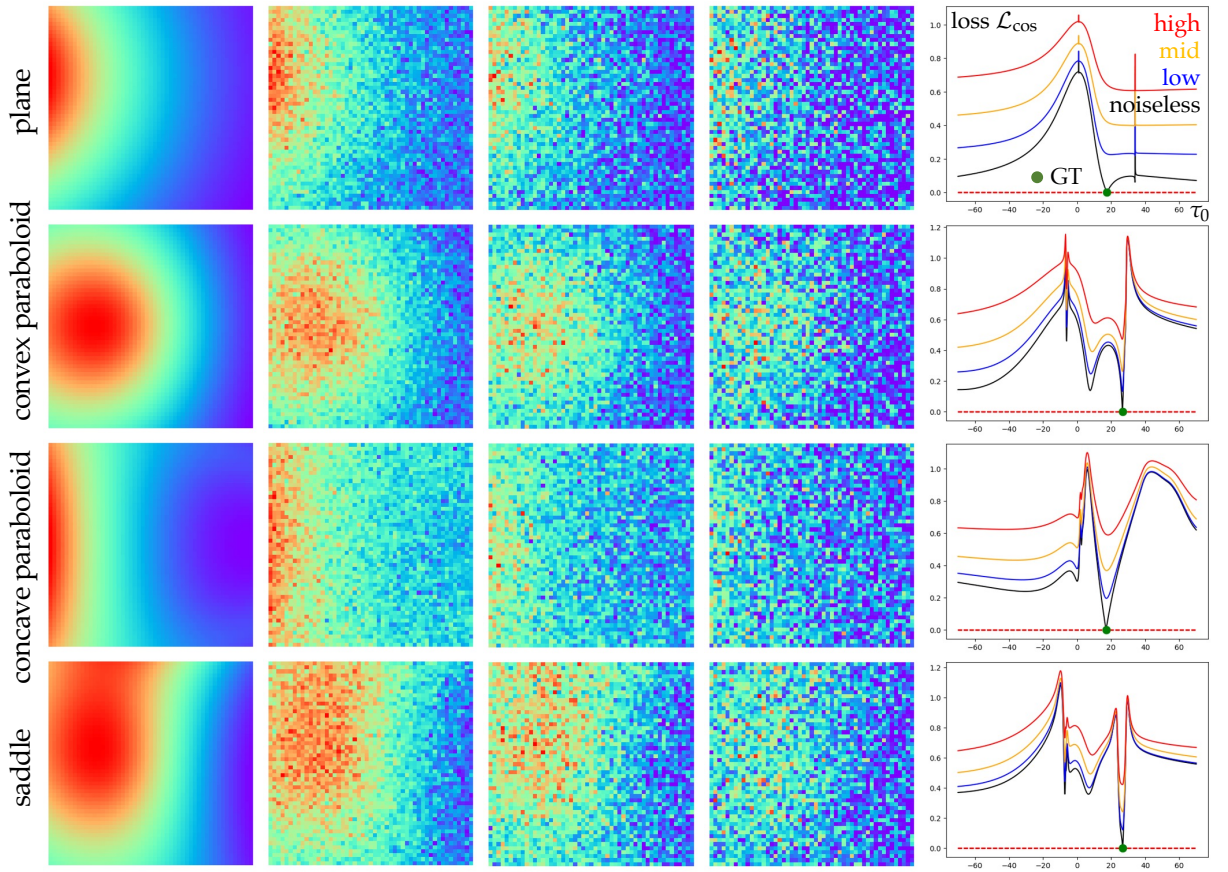


Figure 4.4: **Simulation results on the uniqueness of τ_0 using radiometric measurements.** We consider specular objects of general quadric shapes (shown as depth maps in (a)) and compute (b) image irradiances of specularities observed by a translating camera. (c) The blue and magenta curves both correspond to the polynomial (Equation (4.13)) formulated using one pair of image irradinaces. Multiple zero-crossing points can be found for each polynomial, especially for non-planar surfaces, but many of them are negative and infeasible, which suggests that we have already reduced the ambiguity of τ_0 from infinity to a sparse set of feasible solutions. When considering both polynomials, we see that they usually do not share the same roots except for the ground truth τ_0. The observation suggests that as we include more radiometric measurements, we can further reduce the feasible set of τ_0. (d) Eventually, by considering all irradiances, we obtain a unique solution of τ_0 for all the shapes we simulated.



(a) noiseless irradiances (b) low noise (c) mid noise (d) high noise (e) loss curve comparisons

Figure 4.5: **Simulation results on demonstrating the robustness of the optimization.** We add extensive Gaussian noise to the (a) noiseless radiometric measurements, with signal-to-noise ratios (SNRs) being (b) 10dB, (c) 7dB, and (d) 5dB, respectively. We then run our optimization algorithm using radiometric measurements of all 51×51 specularities. (e) Loss plots show that although the orginal radiometric signal gets corrupted extensively, the global optimum of our formulated optimization remains robustly aligned with the ground truth.

Chapter 5

Defocus Map Estimation and Deblurring from a Single Dual-Pixel Image

Modern DSLR and mirrorless cameras feature large-aperture lenses that allow collecting more light, but also introduce *defocus* blur, meaning that objects in images appear blurred by an amount proportional to their distance from the focal plane. A simple way to reduce defocus blur is to *stop down*, i.e., shrink the aperture. However, this also reduces the amount of light reaching the sensor, making the image noisier. Moreover, stopping down is impossible on fixed-aperture cameras, such as those in most smartphones. More sophisticated techniques fall into two categories. First are techniques that add extra hardware (e.g., coded apertures [101], specialized lenses [43, 102]), and thus are impractical to deploy at large scale or across already available cameras. Second are *focus stacking* techniques [169] that capture multiple images at different focus distances, and fuse them into an all-in-focus image. These techniques require long capture times, and thus are applicable only to static scenes.

Ideally, defocus blur removal should be done using data from a single capture. Unfortunately, in conventional cameras, this task is fundamentally ill-posed: a captured image may have no high-frequency content because either the latent all-in-focus image lacks such frequencies, or they are removed by defocus blur. Knowing the *defocus map*, i.e., the spatially-varying amount of defocus blur, can help simplify blur removal. However, determining the defocus map from a single image is closely-related to monocular

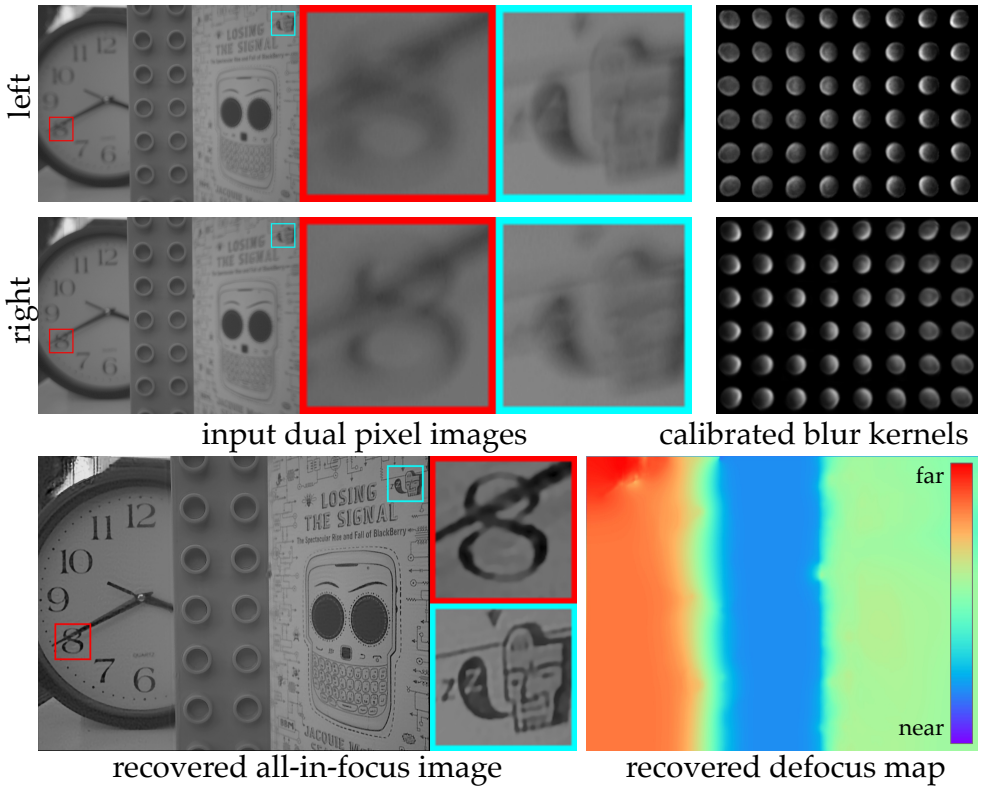


Figure 5.1: Given left and right dual-pixel (DP) images and corresponding spatially-varying blur kernels, our method jointly estimates an all-in-focus image and defocus map.

depth estimation, which is a challenging problem in its own right. Even if the defocus map were known, recovering an all-in-focus image is still an ill-posed problem, as it requires hallucinating the missing high frequency content.

Dual-pixel (DP) sensors are a recent innovation that makes it easier to solve both the defocus map estimation and defocus blur removal problems, with data from a single capture. Camera manufacturers have introduced such sensors to many DSLR and smartphone cameras to improve autofocus [2, 74]. Each pixel on a DP sensor is split into two halves, each capturing light from half of the main lens' aperture, yielding two sub-images per exposure (Figure 5.1). These can be thought of as a two-sample lightfield [129], and their sum is equivalent to the image captured by a regular sensor. The two sub-images have different half-aperture-shaped defocus blur kernels; these are additionally spatially-varying due to optical imperfections such as vignetting or field curvature in lenses, especially for cheap smartphone lenses.

In this chapter, we present a method to simultaneously recover the defocus map and all-in-focus image from a single DP capture. Specifically, we perform a one-time calibration to determine the spatially-varying blur kernels for the left and right DP images. Then, given a single DP image, we optimize a *multiplane image* (MPI) representation [170, 197] to best explain the observed DP images using the calibrated blur kernels. An MPI is a layered representation that accurately models occlusions, and can be used to render both defocused and all-in-focus images, as well as produce a defocus map. As solving for the MPI from two DP images is under-constrained, we introduce additional priors and show their effectiveness via ablation studies. Further, we show that in the presence of image noise, standard optimization has a bias towards underestimating the amount of defocus blur, and we introduce a bias correction term. Our method does not require large amounts of training data, save for a one-time calibration, and outperforms prior art on both defocus map estimation and blur removal, when tested on images captured using a consumer smartphone camera.

5.1 Related Work

Depth estimation. Multi-view depth estimation is a well-posed and extensively studied problem [65, 157]. By contrast, single-view, or *monocular*, depth estimation is ill-posed. Early techniques attempting to recover depth from a single image typically relied on additional cues, such as silhouettes, shading, texture, vanishing points, or data-driven supervision [11, 18, 25, 40, 63, 75, 76, 95, 99, 107, 147, 156, 162]. The use of deep neural networks trained on large RGBD datasets [46, 52, 106, 111, 152, 164] significantly improved the performance of data-driven approaches, motivating approaches that use synthetic data [10, 61, 118, 126, 199], self-supervised training [54, 57, 58, 84, 116, 196], or multiple data sources [47, 148]. Despite these advances, producing high-quality depth from a single image remains difficult, due to the inherent ambiguities of monocular depth estimation.

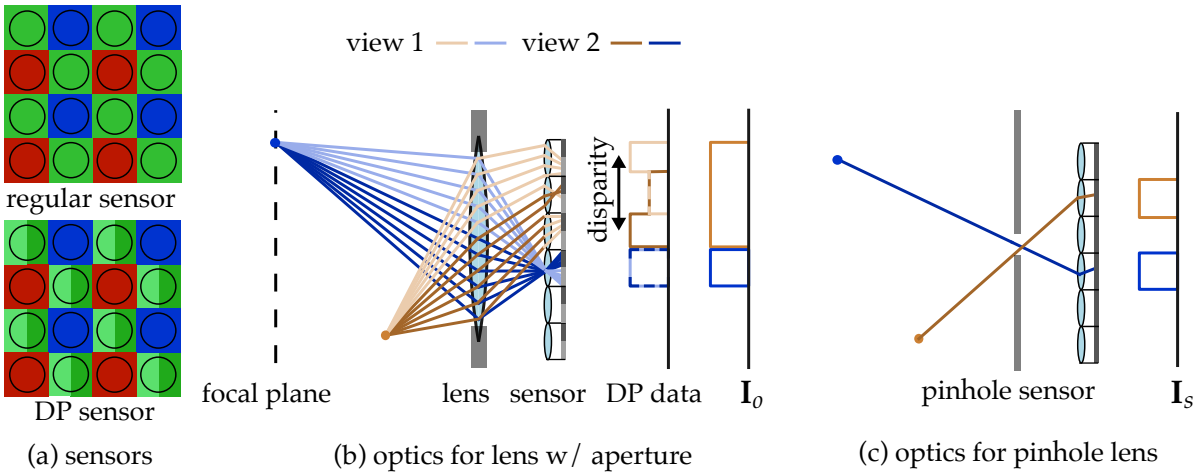
Recent works have shown that DP data can improve monocular depth quality, by resolving some of these ambiguities. Wadhwa *et al*. [181] applied classical stereo matching methods to DP views to compute depth. Punnappurath *et al*. [144] showed that explicitly modeling defocus blur during stereo matching can improve depth quality. However, they assume that the defocus blur is spatially invariant and symmetric between the

left and right DP images, which is not true in real smartphone cameras. Depth estimation with DP images has also been used as part of reflection removal algorithms [143]. Garg *et al*. [53] and Zhang *et al*. [192] trained neural networks to output depth from DP images, using a captured dataset of thousands of DP images and ground truth depth maps [8]. The resulting performance improvements come at a significant data collection cost.

Focus or defocus has been used as a cue for monocular depth estimation prior to these DP works. Depth from defocus techniques [48, 141, 173, 185] use two differently-focused images with the same viewpoint, whereas depth from focus techniques use a dense focal stack [59, 68, 169]. Other monocular depth estimation techniques use defocus cues as supervision for training depth estimation networks [167], use a coded aperture to estimate depth from one [101, 178, 193] or two captures [194], or estimate a defocus map using synthetic data [100]. Lastly, some binocular stereo approaches also explicitly account for defocus blur [37, 105]; compared to depth estimation from DP images, these approaches assume different focus distances for the two views.

Defocus deblurring. Besides depth estimation, measuring and removing defocus blur is often desirable to produce sharp all-in-focus images. Defocus deblurring techniques usually estimate either a depth map or an equivalent defocus map as a first processing stage [42, 86, 137, 163]. Some techniques modify the camera hardware to facilitate this stage. Examples include inserting patterned occluders in the camera aperture to make defocus scale selection easier [101, 178, 193, 194]; or sweeping through multiple focal settings within the exposure to make defocus blur spatially uniform [125]. Once a defocus map is available, a second deblurring stage employs non-blind deconvolution methods [51, 96, 101, 120, 184, 191] to remove the defocus blur.

Deep learning has been successfully used for defocus deblurring as well. Lee *et al*. [100] train neural networks to regress to defocus maps, that are then used to deblur. Abuolaim and Brown [1] extend this approach to DP data, and train a neural network to directly regress from DP images to all-in-focus images. Their method relies on a dataset of pairs of wide and narrow aperture images captured with a DSLR, and may not generalize to images captured on smartphone cameras, which have very different optical characteristics. Such a dataset is impossible to collect on smartphone cameras with fixed aperture lenses. In contrast to these prior works, our method does not require difficult-to-capture large datasets. Instead, it uses an accurate model of the defocus blur characteristics of DP data, and simultaneously solves for a defocus map and an all-in-



(a) sensors (b) optics for lens w/ aperture (c) optics for pinhole lens

Figure 5.2: **Dual-pixel (DP) image formation.** (a) A regular sensor and a DP sensor where each green pixel is split into two halves. (b) For a finite aperture lens, an in-focus scene point produces overlapping DP images, whereas an out-of-focus point produces shifted DP images. Adding the two DP images yields the image that would have been captured by a regular sensor. In (c), we show the corresponding pinhole camera where all scene content is in focus. Ignoring occlusions, images in (b) can be generated from the image in (c) by applying a depth-dependent blur.

focus image.

5.2 Dual-Pixel Image Formation

We begin by describing image formation for a regular and a dual-pixel (DP) sensor, to relate the defocus map and the all-in-focus image to the captured image. For this, we consider a camera imaging a *diffuse* scene with two points, only one of which is in focus (Figure 5.2(b)). Rays emanating from the in-focus point (blue) converge on a single pixel, creating a sharp image. By contrast, rays from the out-of-focus point (brown) fail to converge, creating a blurred image.

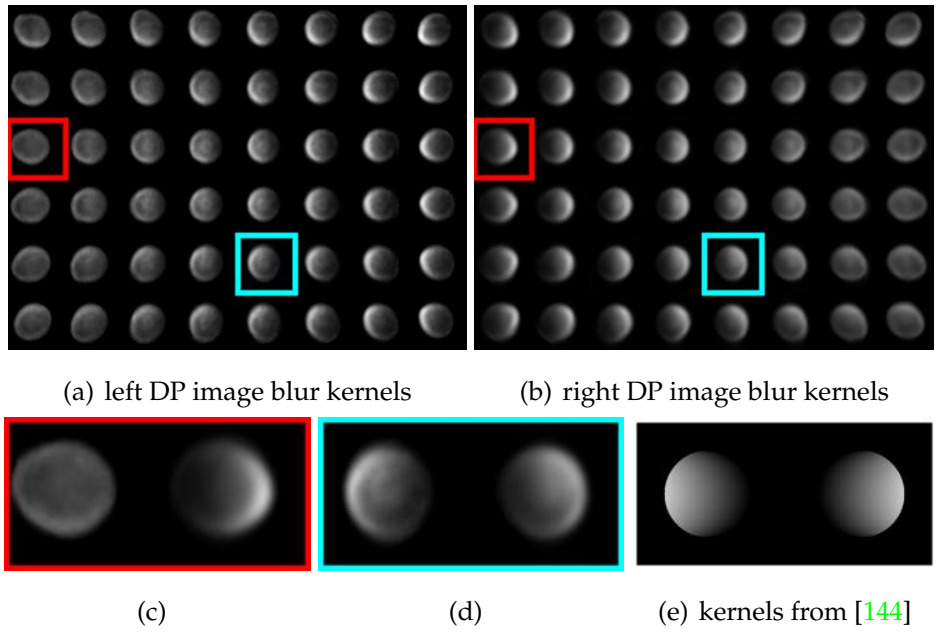
If we consider a lens with an infinitesimally-small aperture (i.e., a pinhole camera), only rays that pass through its center strike the sensor, and create a sharp all-in-focus image $\mathbf{I}_s$ (Figure 5.2(c)). Under the thin lens model, the blurred image $\mathbf{I}_o$ of the out-of-focus point equals blurring $\mathbf{I}_s$ with a depth-dependent kernel $\boldsymbol{k}_d$, shaped as a d-scaled version of the aperture—typically a circular disc of radius $d = A + {}^{B}/_{Z}$, where Z is the

point depth, and A and B are lens-dependent constants [53]. Therefore, the per-pixel signed kernel radius d, termed the *defocus map* $\mathbf{D}$, is a linear function of inverse depth, thus a proxy for the depth map. Given the defocus map $\mathbf{D}$, and ignoring occlusions, the sharp image $\mathbf{I}_s$ can be recovered from the captured image $\mathbf{I}_o$ using non-blind deconvolution. In practice, recovering either the defocus map $\mathbf{D}$ or the sharp image $\mathbf{I}_s$ from a single image $\mathbf{I}_o$ is ill-posed, as multiple $(\mathbf{I}_s, \mathbf{D})$ combinations produce the same image $\mathbf{I}_o$. Even when the defocus map $\mathbf{D}$ is known, determining the sharp image $\mathbf{I}_s$ is still ill-posed, as blurring irreversibly removes image frequencies.

DP sensors make it easier to estimate the defocus map. In DP sensors (Figure 5.2(a)), each pixel is split into two halves, each collecting light from the corresponding half of the lens aperture (Figure 5.2(b)). Adding the two half-pixel, or DP, images $\mathbf{I}_o^l$ and $\mathbf{I}_o^r$ produces an image equivalent to that captured by a regular sensor, i.e., $\mathbf{I}_o = \mathbf{I}_o^l + \mathbf{I}_o^r$. Furthermore, DP images are identical for an in-focus scene point, and shifted versions of each other for an out-of-focus point. The amount of shift, termed *DP disparity*, is proportional to the blur size, and thus provides an alternative for defocus map estimation. In addition to facilitating the estimation of the defocus map $\mathbf{D}$, having two DP images instead of a single image provides additional constraints for recovering the underlying sharp image $\mathbf{I}_s$. Utilizing these constraints requires knowing the blur kernel shapes for the two DP images.

Blur kernel calibration. As real lenses have spatially-varying kernels, we calibrate an 8×6 grid of kernels. To do this, we fix the focus distance, capture a regular grid of circular discs on a monitor screen, and solve for blur kernels for left and right images independently using a method similar to Mannan and Langer [117]. When solving for kernels, we assume that they are normalized to sum to one, and calibrate separately for vignetting: we average left and right images from six captures of a white diffuser, using the same focus distance as above, to produce left and right vignetting patterns $\mathbf{W}_l$ and $\mathbf{W}_r$. We refer to Section 5.4.2 for more details.

We show the calibrated blur kernels in Figure 5.3. We note that these kernels deviate significantly from parametric models derived by extending the thin lens model to DP image formation [144]. In particular, the calibrated kernels are spatially-varying, not circular, and not symmetric.



(a) left DP image blur kernels (b) right DP image blur kernels

(c) (d) (e) kernels from [144]

Figure 5.3: Calibrated blur kernels (a) and (b) for the left and right DP images. (c) and (d) show example pairs of left and right kernels marked in red and cyan. Compared to the parametric kernels (e) from [144], calibrated kernels are spatially-varying, not circular, and not left-right symmetric.

5.3 Proposed Method

The inputs to our method are two single-channel DP images, and calibrated left and right blur kernels. We correct for vignetting using W_l and W_r, and denote the two vignetting-corrected DP images as $\mathbf{I}_o^l$ and $\mathbf{I}_o^r$, and their corresponding blur kernels at a certain defocus size d as $\boldsymbol{k}_d^l$ and $\boldsymbol{k}_d^r$, respectively. We assume that blur kernels at a defocus size d' can be obtained by scaling by a factor d'/d [144, 194]. Our goal is to optimize for the multiplane image (MPI) representation that best explains the observed data, and use it to recover the latent all-in-focus image $\hat{\mathbf{I}}_s$ and defocus map $\hat{\mathbf{D}}$. We first introduce the MPI representation, and show how to render defocused images from it. We then formulate an MPI optimization problem, and detail its loss function.

5.3.1 Multiplane Image (MPI) Representation

We model the scene using the MPI representation, previously used primarily for view synthesis [177, 197]. MPIs discretize the 3D space into N fronto-parallel planes at fixed

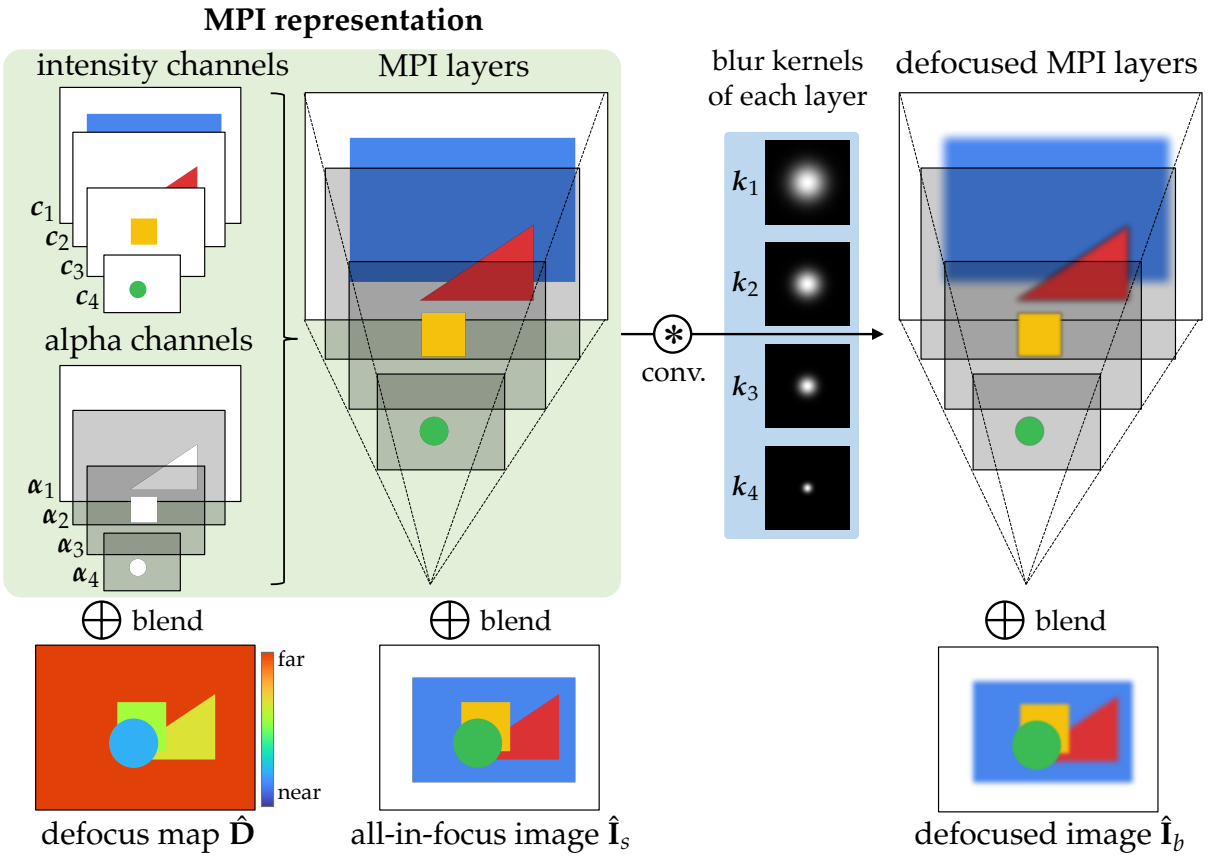


Figure 5.4: **Multiplane image (MPI) representation.** An MPI consists of discrete fronto-parallel planes where each plane contains intensity data and an alpha channel. We use it to recover the defocus map, the all-in-focus image, and render a defocused image according to a given blur kernel.

depths (Figure 5.4). We select depths corresponding to linearly-changing defocus blur sizes $[d_1, \ldots, d_N]$. Each MPI plane is an intensity-alpha image of the in-focus scene that consists of an intensity channel $\boldsymbol{c}_i$ and an alpha channel $\boldsymbol{\alpha}_i$.

All-in-focus image compositing. Given an MPI, we composite the sharp image using the *over* operator [114]: we sum all layers weighted by the transmittance of each layer $\boldsymbol{t}_i$,

$$\hat{\mathbf{I}}_s = \sum_{i=1}^{N} \boldsymbol{t}_i \boldsymbol{c}_i = \sum_{i=1}^{N} \left[\boldsymbol{c}_i \boldsymbol{\alpha}_i \prod_{j=i+1}^{N} (1 - \boldsymbol{\alpha}_j) \right] . \tag{5.1}$$

Defocus map rendering. We can synthesize a continuous-valued defocus map $\hat{\mathbf{D}}$ in a similar way as discussed by Tucker and Snavely [177], by replacing all pixel intensities

in Equation (5.1) with the defocus blur size d_i of that layer:

$$\hat{\mathbf{D}} = \sum_{i=1}^{N} \left[d_i \boldsymbol{\alpha}_i \prod_{j=i+1}^{N} (1 - \boldsymbol{\alpha}_i) \right] . \tag{5.2}$$

Defocused image rendering. Given the left and right blur kernels $\boldsymbol{k}_{d_i}^{\{l,r\}}$ for each layer, we render defocused images by convolving each layer with its corresponding kernel, then compositing the blurred layers as in Equation (5.1):

$$\hat{\mathbf{I}}_b^{\{l,r\}} = \sum_{i=1}^{N} \left[\left(\boldsymbol{k}_{d_i}^{\{l,r\}} * (\boldsymbol{c}_i \boldsymbol{\alpha}_i) \right) \odot \prod_{j=i+1}^{N} \left(\mathbf{1} - \boldsymbol{k}_{d_j}^{\{l,r\}} * \boldsymbol{\alpha}_j \right) \right] , \tag{5.3}$$

where $*$ denotes convolution. In practice, we scale the calibrated spatially-varying left and right kernels by the defocus size d_i, and apply the scaled spatially-varying blur to each intensity-alpha image $\boldsymbol{c}_i \boldsymbol{\alpha}_i$. We note that we render left and right views from a single MPI, but with different kernels.

5.3.2 Effect of Gaussian Noise on Defocus Estimation

Using Equation (5.3), we can optimize for the MPI that minimizes the L_2 loss $\left\| \hat{\mathbf{I}}_b^{\{l,r\}} - \mathbf{I}_o^{\{l,r\}} \right\|_2^2$ between rendered images $\hat{\mathbf{I}}_b^{\{l,r\}}$ and observed DP images $\mathbf{I}_o^{\{l,r\}}$. Here we show that, in the presence of noise, this optimization is biased toward smaller defocus sizes, and we correct for this bias.

Assuming additive white Gaussian noise $\mathbf{N}^{\{l,r\}}$ distributed as $\mathcal{N}(0, \sigma^2)$, we can model DP images as:

$$\mathbf{I}_o^{\{l,r\}} = \mathbf{I}_b^{\{l,r\}} + \mathbf{N}^{\{l,r\}} , \tag{5.4}$$

where $\mathbf{I}_b^{\{l,r\}}$ are the latent noise-free images. For simplicity, we assume for now that all scene content lies on a single fronto-parallel plane with ground truth defocus size $d^\star$. Then, using frequency domain analysis similar to Zhou *et al.* [194], we prove in Appendix D.1 that for a defocus size hypothesis d_i, the expected negative log-energy function corresponding to the MAP estimate of the MPI is:

$$\begin{aligned} E\left(d_i | \boldsymbol{K}_{d^\star}^{\{l,r\}}, \sigma\right) = & \sum_f C_1\left(\boldsymbol{K}_{d_i}^{\{l,r\}}, \sigma, \boldsymbol{\Phi}\right) \left| \boldsymbol{K}_{d^\star}^l \boldsymbol{K}_{d_i}^r - \boldsymbol{K}_{d^\star}^r \boldsymbol{K}_{d_i}^l \right|^2 + \\ & \sigma^2 \sum_f \left[\frac{|\boldsymbol{K}_{d^\star}^l|^2 + |\boldsymbol{K}_{d^\star}^r|^2 + \sigma^2 |\boldsymbol{\Phi}|^2}{|\boldsymbol{K}_{d_i}^l|^2 + |\boldsymbol{K}_{d_i}^r|^2 + \sigma^2 |\boldsymbol{\Phi}|^2} \right] + C_2(\sigma) , \end{aligned} \tag{5.5}$$

where $\boldsymbol{K}_{d_i}^{\{l,r\}}$ and $\boldsymbol{K}_{d^\star}^{\{l,r\}}$ are the Fourier transforms of kernels $\boldsymbol{k}_{d_i}^{\{l,r\}}$ and $\boldsymbol{k}_{d^\star}^{\{l,r\}}$ respectively, $\boldsymbol{\Phi}$ is the inverse spectral power distribution of natural images, and the summation is over all frequencies. We would expect the loss to be minimized when $d_i = d^\star$. The first term measures the inconsistency between the hypothesized blur kernel d_i and the true kernel $d^\star$, and is indeed minimized when $d_i = d^\star$. However, the second term depends on the noise variance and decreases as $|d_i|$ decreases. This is because, for a normalized blur kernel ($\left\|\boldsymbol{k}_{d_i}^{\{l,r\}}\right\|_1 = 1$), as the defocus kernel size $|d_i|$ decreases, its power spectrum $\left\|\boldsymbol{K}_{d_i}^{\{l,r\}}\right\|_2$ increases. This suggests that white Gaussian noise in input images results in a bias towards smaller blur kernels. To account for this bias, we subtract an approximation of the second term, which we call the *bias correction term*, from the optimization loss:

$$\mathcal{B}\left(d_i|\boldsymbol{K}_{d^\star}^{\{l,r\}},\sigma\right) \approx \sigma^2 \sum_f \frac{\sigma^2 \left|\boldsymbol{\Phi}\right|^2}{\left|\boldsymbol{K}_{d_i}^{l}\right|^2 + \left|\boldsymbol{K}_{d_i}^{r}\right|^2 + \sigma^2 \left|\boldsymbol{\Phi}\right|^2}\,. \tag{5.6}$$

We ignore the terms containing ground truth $d^\star$, as they are significant only when $d^\star$ is itself small, i.e., the bias favors the true kernels in that case. In an MPI with multiple layers associated with defocus sizes $[d_1, \ldots, d_N]$, we subtract per-layer constants $\mathcal{B}\left(d_i\right)$ computed using Equation (5.6).

We note that we use a Gaussian noise model to make analysis tractable, but captured images have mixed Poisson-Gaussian noise [66]. In practice, we found it beneficial to additionally denoise the input images using burst denoising [67]. However, there is residual noise even after denoising, and we show in Section 5.4.4 that our bias correction term still improves performance. An interesting future research direction is using a more accurate noise model to derive a better bias estimate and remove the need for any denoising.

5.3.3 MPI Optimization

We seek to recover an MPI $\{\boldsymbol{c}_i, \boldsymbol{\alpha}_i\}, i \in [1, \ldots, N]$ such that defocused images rendered from it using the calibrated blur kernels are close to the input images. But minimizing only a reconstruction loss is insufficient: this task is ill-posed, as there exists an infinite family of MPIs that all exactly reproduce the input images. As is common in defocus deblurring [101], we regularize our optimization (Figure 5.5):

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{data}} + \lambda_2 \mathcal{L}_{\text{aux}} + \lambda_3 \mathcal{L}_{\text{intensity}} + \lambda_4 \mathcal{L}_{\text{alpha}} + \lambda_5 \mathcal{L}_{\text{entropy}}\,. \tag{5.7}$$

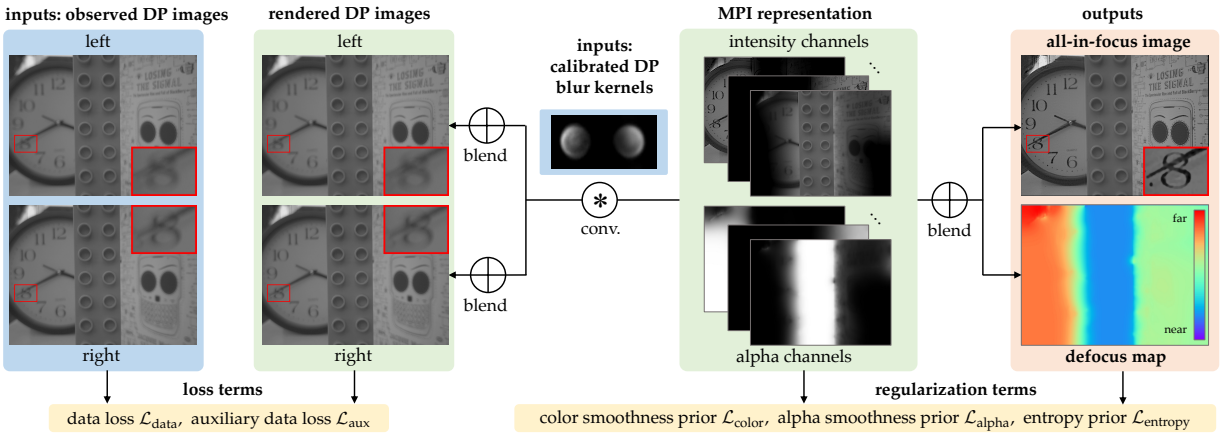


Figure 5.5: **Overview of our proposed method.** We use input left and right DP images to fit a multiplane image (MPI) scene representation, consisting of a set of fronto-parallel layers. Each layer is an intensity-alpha image containing the in-focus scene content at the corresponding depth. The MPI can output the all-in-focus image and the defocus map by blending all layers. It can also render out-of-focus images, by convolving each layer with pre-calibrated blur kernels for the left and right DP views, and then blending. We optimize the MPI by minimizing a regularized loss comparing rendered and input images.

where $\mathcal{L}_{\text{data}}$ is a bias-corrected data term that encourages rendered images to resemble input images, $\mathcal{L}_{\text{aux}}$ is an auxiliary data term applied to each MPI layer, and the remaining are regularization terms. We discuss all loss terms below, and specify the scaling factors λ_1-λ_5 used in our experiments in Section 5.4.3.

Bias-corrected data loss. We consider the Charbonnier [36] loss function $\ell(x) = \sqrt{x^2/\gamma^2 + 1}$, and define a bias-corrected version as $\ell_{\mathcal{B}}(x, \mathcal{B}) = \sqrt{(x^2 - \mathcal{B})/\gamma^2 + 1}$, where we choose the scale parameter $\gamma = 0.1$ [17]. We use this loss function to form a data loss penalizing the difference between left and right input and rendered images as:

$$\mathcal{L}_{\text{data}} = \sum_{x,y} \ell_{\mathcal{B}}\left(\hat{\mathbf{I}}_b^{\{l,r\}}(x,y) - \mathbf{I}_o^{\{l,r\}}(x,y), \mathcal{B}_{\text{all}}^{\{l,r\}}\right), \tag{5.8}$$

$$\mathcal{B}_{\text{all}}^{\{l,r\}} = \sum_{i=1}^{N} \left[\boldsymbol{k}_{d_i}^{\{l,r\}} * \boldsymbol{\alpha}_i \prod_{j=i+1}^{N} \left(1 - \boldsymbol{k}_{d_j}^{\{l,r\}} * \boldsymbol{\alpha}_j\right) \right] \mathcal{B}(d_i). \tag{5.9}$$

We compute the total bias correction $\mathcal{B}_{\text{all}}^{\{l,r\}}$ as the sum of all bias correction terms of each layer, weighted by the corresponding defocused transmittance. Equation (5.9) is equivalent to Equation (5.3) where we replace each MPI layer's intensity channel $\boldsymbol{c}_i$

with a constant bias correction value $\mathcal{B}(d_i)$. To compute $\mathcal{B}(d_i)$ from Equation (5.6), we empirically set the variance to $\sigma^2 = 5 \cdot 10^{-5}$, and use a constant inverse spectral power distribution $|\mathbf{\Phi}|^2 = 10^2$, following previous work [172].

Auxiliary data loss. In most real-world scenes, a pixel's scene content should be on a single layer. However, because the compositing operator of Equation (5.3) forms a weighted sum of all layers, $\mathcal{L}_{\text{data}}$ can be small even when scene content is smeared across multiple layers. To discourage this, we introduce a per-layer auxiliary data loss on each layer's intensity weighted by the layer's blurred transmittance:

$$\mathcal{L}_{\text{aux}} = \sum_{x,y,i} \left(\boldsymbol{k}_{d_i}^{\{l,r\}} * \boldsymbol{t}_i(x,y) \right) \odot \ell_{\mathcal{B}} \left(\boldsymbol{k}_{d_i}^{\{l,r\}} * \boldsymbol{c}_i(x,y) - \mathbf{I}_o^{\{l,r\}}(x,y), \mathcal{B}(d_i) \right) , \tag{5.10}$$

where $\odot$ denotes element-wise multiplication. This auxiliary loss resembles the data synthesis loss of Equation (5.8), except that it is applied to each MPI layer separately.

Intensity smoothness. Our first regularization term encourages smoothness for the all-in-focus image and the MPI intensity channels. For an image $\mathbf{I}$ with corresponding edge map $\boldsymbol{E}$, we define an edge-aware smoothness based on total variation $\boldsymbol{V}(\cdot)$, similar to Tucker and Snavely [177]:

$$\boldsymbol{V}_{\boldsymbol{E}}(\mathbf{I}, \boldsymbol{E}) = \ell(\boldsymbol{V}(\mathbf{I})) + (1 - \boldsymbol{E}) \odot \ell(\boldsymbol{V}(\mathbf{I})) , \tag{5.11}$$

where $\ell(\cdot)$ is the Charbonnier loss. We refer to Appendix D.2 for detials on $\boldsymbol{E}$ and $\boldsymbol{V}(\cdot)$. Our smoothness prior on the all-in-focus image and MPI intensity channels is:

$$\mathcal{L}_{\text{intensity}} = \sum_{x,y} \boldsymbol{V}_{\boldsymbol{E}}(\hat{\mathbf{I}}_s, \boldsymbol{E}(\hat{\mathbf{I}}_s)) + \sum_{x,y,i} \boldsymbol{V}_{\boldsymbol{E}}(\boldsymbol{t}_i \boldsymbol{c}_i, \boldsymbol{E}(\boldsymbol{t}_i \boldsymbol{c}_i)) . \tag{5.12}$$

Alpha and transmittance smoothness. We use an additional smoothness regularizer on all alpha channels and transmittances (sharpened by computing their square root), by encouraging edge-aware smoothness according to the total variation of the all-in-focus image:

$$\mathcal{L}_{\text{alpha}} = \sum_{x,y,i} \left[\boldsymbol{V}_{\boldsymbol{E}}(\sqrt{\boldsymbol{\alpha}_i}, \boldsymbol{E}(\hat{\mathbf{I}}_s)) + \boldsymbol{V}_{\boldsymbol{E}}(\sqrt{\boldsymbol{t}_i}, \boldsymbol{E}(\hat{\mathbf{I}}_s)) \right] . \tag{5.13}$$

Alpha and transmittance entropy. The last regularizer is a collision entropy penalty on alpha channels and transmittances. Collision entropy, defined for a vector $\boldsymbol{x}$ as $S(\boldsymbol{x}) = -\log \|x\|_2^2 / \|x\|_1^2$, is a special case of Renyi entropy [149], and we empirically found it to be better than Shannon entropy for our problem. Minimizing collision entropy encourages sparsity: $S(\boldsymbol{x})$ is minimum when all but one elements of $\boldsymbol{x}$ are 0, which in our case

encourages scene content to concentrate on a single MPI layer, rather than spread across multiple layers. Our entropy loss is:

$$\mathcal{L}_{\text{entropy}} = \sum_{x,y} S\left(\left[\sqrt{\boldsymbol{\alpha}_2}\,(x,y), \ldots, \sqrt{\boldsymbol{\alpha}_N}\,(x,y)\right]^{\mathrm{T}}\right) + \sum_{x,y} S\left(\left[\sqrt{\boldsymbol{t}_1}\,(x,y), \ldots, \sqrt{\boldsymbol{t}_N}\,(x,y)\right]^{\mathrm{T}}\right). \tag{5.14}$$

We extract the alpha channels and transmittances of each pixel (x, y) from all MPI layers, compute their square root for sharpening, compute a per-pixel entropy, and average these entropies across all pixels. When computing entropy on alpha channels, we skip the farthest MPI layer, because we assume that all scene content ends at the farthest layer, and thus force this layer to be opaque ($\boldsymbol{\alpha}_1 = \mathbf{1}$).

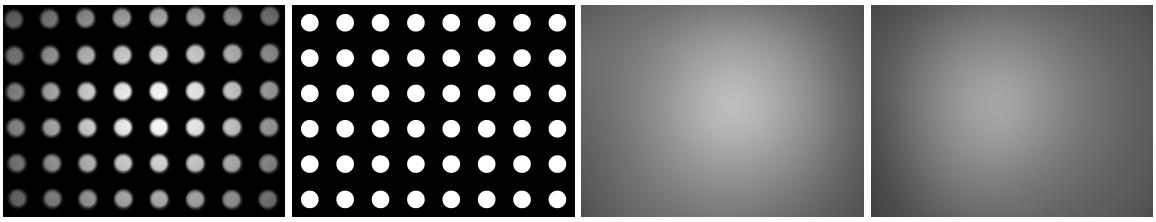
5.4 Experiments

We capture a new dataset, and use it to perform qualitative and quantitative comparisons with other state of the art defocus deblurring and defocus map estimation methods.

5.4.1 Data Collection

Even though DP sensors are common, to the best of our knowledge, only two camera manufacturers provide an API to read DP images—Google and Canon. However, Canon's proprietary software applies an unknown scene-dependent transform to DP data. Unlike supervised learning-based methods [1] that can learn to account for this transform, our loss function requires raw sensor data. Hence, we collect data using the Google Pixel 4 smartphone, which allows access to the raw DP data [44]. We capture a total of 17 scenes, both indoors and outdoors. Similar to Garg *et al*. [53], we centrally crop the DP images to 1008×1344.

"Ground truth" generation. The Pixel 4 captures DP data only in the green channel. To compute ground truth, we capture a focus stack with 36 slices sampled uniformly in diopter space, where the closest focus distance corresponds to the distance we calibrate for, 13.7 cm, and the farthest to infinity. Following prior work [144], we use the commercial Helicon Focus software [72] to process the stacks and generate ground truth sharp images and defocus maps, and we manually correct holes in the generated defocus maps. Still, there are image regions that are difficult to manually inpaint, e.g.,



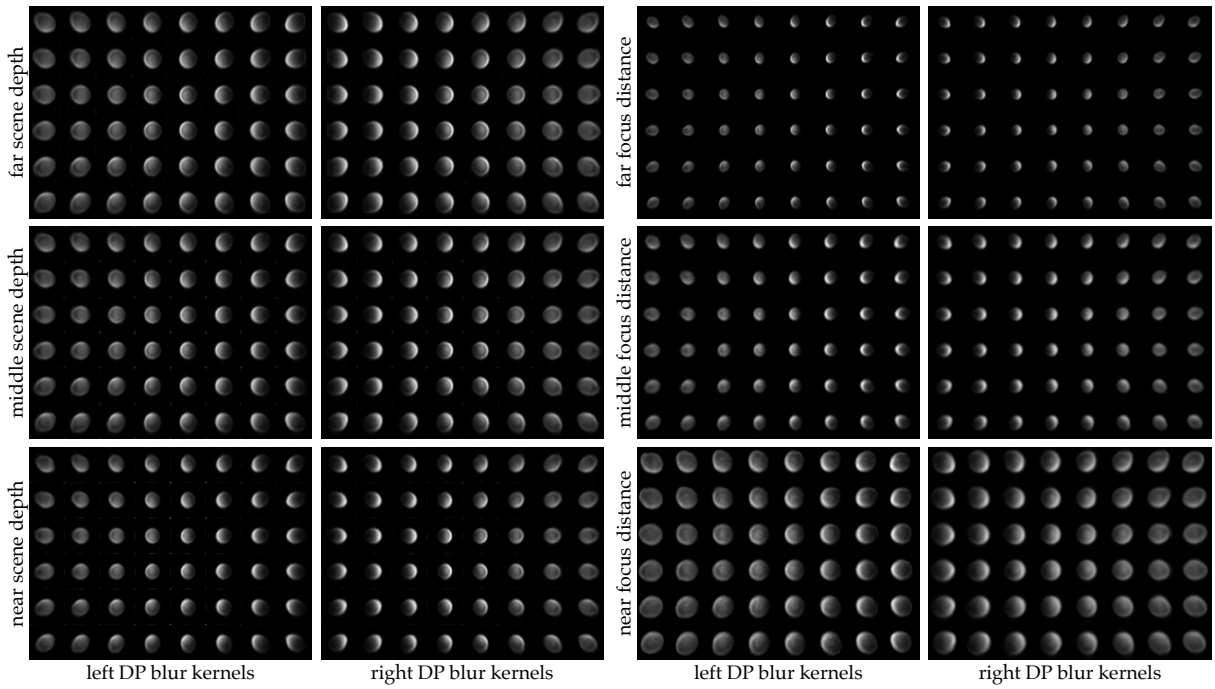
(a) captured image (left + right) (b) calibration pattern (c) left vignetting (d) right vignetting

Figure 5.6: **Blur kernel calibration.** We show (a) the captured DP image (left + right) of (b) the calibration pattern that is used to calibrate the blur kernels. In addition to blur kernels, we calibrate for different vignetting in (c) left and (d) right DP images by capturing a white sheet through a diffuser.

near occlusion boundaries or curved surfaces. We ignore such regions when computing quantitative metrics. There may be a small *shift* between the ground truth all-in-focus image and the all-in-focus image from the deblurring algorithms we evaluate. This is because one can apply an arbitrary transform to the blur kernels and an inverse transform to the recovered all-in-focus image to yield the same blurred image. We determine this shift for each algorithm by using OpenCV to align the ground-truth all-in-focus image with the all-in-focus image from the algorithm via an affine transform for a single specific scene, and then using that transform to align all images before computing the metrics for all-in-focus images. We also crop a small border of 8 pixels before computing the metrics as it may contain invalid pixels after alignment. We note that we only use the generated ground truth for quantitative evaluations, and not for training.

5.4.2 Blur Kernel Calibration

We provide more information about our calibration procedure for the left and right blur kernels used as input to our method. As shown in Figure 5.6, we image a regular grid of circular discs on a monitor screen at a distance of ~ 45 cm from the camera. We apply global thresholding and binarize the captured image, perform connected component analysis to identify the individual discs and their centers, and generate and align the binary sharp image $\mathbf{M}$ with the known calibration pattern by solving for a homography between the calibration target disc centers and the detected centers. In order to apply radiometric correction, we capture all-white and all-black images displayed on the



(a) DP blur kernels with respect to scene depth (b) DP blur kernels with respect to focus distance

Figure 5.7: **DP blur kernels with respect to scene depth and focus distance.** We choose focus settings such that all scene contents are at or behind the focus plane, and calibrate for blur kernels either (a) with the same focus settings but at different depths, or (b) at the same depth but with various focus distances.

same screen, denoted as $\mathbf{I}_w$ and $\mathbf{I}_b$, respectively. We then generate the grayscale latent sharp all-in-focus image as $\mathbf{I}_s = \mathbf{M} \odot \mathbf{I}_w + (\mathbf{1} - \mathbf{M}) \odot \mathbf{I}_b$, where $\odot$ represents pixel-wise multiplication. Once we have the aligned latent image and the captured image, we can solve for spatially-varying blur kernels using the optimization proposed by Mannan and Langer [117]. Specifically, we solve for an 8×6 grid of kernels corresponding to 1344×1008 central field of view.

In addition to the blur kernels, we calibrate for different vignetting in left and right DP images. Specifically, for the same focus distance as above, we capture six images of a white sheet through a diffuser. We then average all left and right images individually to obtain the left and right vignetting patterns $\mathbf{W}_l$ and $\mathbf{W}_r$, respectively.

Next, we explore how DP blur kernels change with respect to scene depth and focus distance (Figure 5.7). As observed by Tang and Kutulakos [172], we find that kernels behave differently on the opposite sides of the focus plane. Therefore we choose fo-

cus settings such that all scene contents are at or behind the focus plane for all of our experiments, including this kernel analysis. We observe that DP blur kernels are approximately resized versions of each other as the scene depth or focus distance changes, similar to the expected behavior for blur kernels in a regular image sensor.

5.4.3 Implementation Details

We discuss implementation details of our method.

Data normalization. Before running the optimization, we first compute an intensity scaling factor $s = {}^{0.5}/_{\text{mean}}\left(\mathbf{I}_o^{\{l,r\}}\right)$, and normalize the inputs $\bar{\mathbf{I}}_o^{\{l,r\}} = s\mathbf{I}_o^{\{l,r\}}$ to account for global intensity variations. After optimization, we undo the normalization by dividing the all-in-focus image by s.

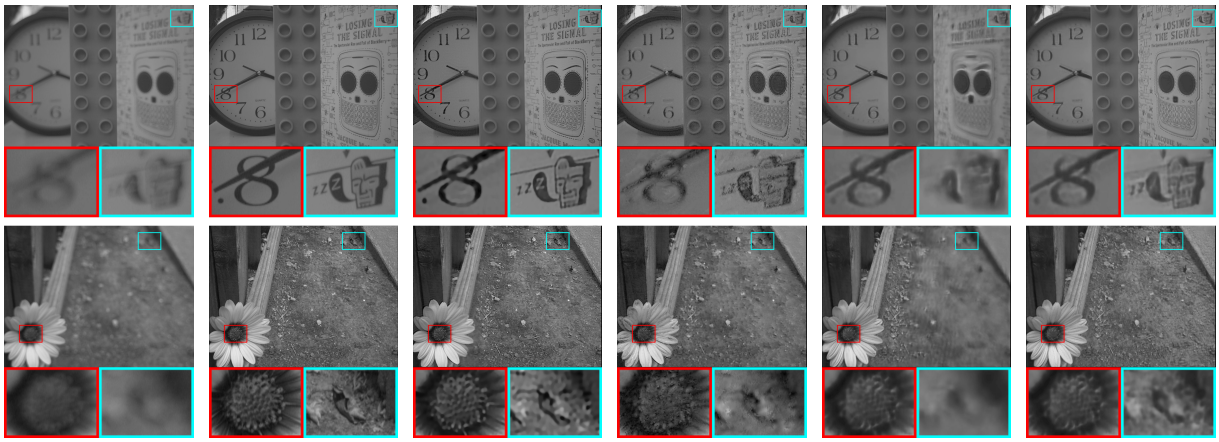
MPI Optimization. We use $N = 12$ MPI layers for all scenes in our dataset. We manually determine the kernel sizes of the front and back layers, and evenly distribute layers in diopter space. As mentioned in Section 5.4.2, we choose focus settings such that all scene contents are at or behind the focus plane. Therefore, the kernel size of the front MPI layer is usually set to a small positive number, e.g., in the range of 1×1 to 3×3, to mimic a 2D delta function, while the kernel size corresponding to the back MPI layer is set to a large enough value, e.g., in the range of 57×57 to 61×61, to represent blur kernels at infinity.

Recall our optimization loss (Equation (5.7)) is $\mathcal{L} = \lambda_1\mathcal{L}_{\text{data}} + \lambda_2\mathcal{L}_{\text{aux}} + \lambda_3\mathcal{L}_{\text{intensity}} + \lambda_4\mathcal{L}_{\text{alpha}} + \lambda_5\mathcal{L}_{\text{entropy}}$. We set the same weight on $\mathcal{L}_{\text{data}}$ and $\mathcal{L}_{\text{aux}}$: $\lambda_1 = \lambda_2 = 2.5 \cdot 10^4$. For most scenes, $\lambda_3 = 30$, $\lambda_4 = 7.5 \cdot 10^4$, and $\lambda_5 = 12$. We set higher weights on the regularization terms $\mathcal{L}_{\text{intensity}}$, $\mathcal{L}_{\text{alpha}}$, and $\mathcal{L}_{\text{entropy}}$ for scenes with less texture, e.g., data from Abuolaim and Brown [1].

Each optimization runs for 10,000 iterations with Adam optimizer [90], and takes 2 hours on an Nvidia Titan RTX GPU. We gradually decrease the global learning rate from 0.3 to 0.1 with exponential decay.

5.4.4 Results

We evaluate our method on both defocus deblurring and depth-from-defocus tasks, and compare to state-of-the-art methods for defocus deblurring (DPDNet [1], Wiener deconvolution [172, 194]) and defocus map estimation (DP stereo matching [181], supervised



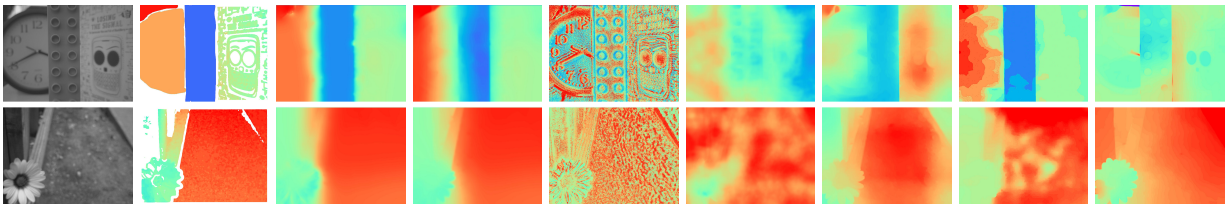
(a) input image (b) GT all-in-focus image (c) ours (d) Wiener deconv. [194] (e) DPDNet [1] (Orig. Input) (f) DPDNet [1]

Figure 5.8: **Qualitative comparisons of various defocus deblurring methods.** We show (a) input images as the average of two DP views, (b) ground truth all-in-focus images computed from focus stacks, and recovered all-in-focus images (c) from our method and other methods (d)-(f). We improve the accuracy of (e) DPDNet trained on Canon data by providing (f) vignetting-corrected images. Our method performs the best in recovering high-frequency details and presents fewer artifacts.

learning from DP views [53], DP defocus estimation based on kernel symmetry [144], Wiener deconvolution [172, 194], DMENet [100]). For methods that take a single image as input, we use the average of the left and right DP images. We also provide both the original and vignetting corrected DP images as inputs, and report the best result. We show quantitative results in Table 5.1 and qualitative results in Figures 5.8 and 5.9. For the defocus map, we use the affine-invariant metrics from Garg *et al*. [53]. Our method achieves the best quantitative results on both tasks.

Defocus deblurring results. Despite the large amount of blur in the input DP images, our method produces deblurred results with high-frequency details that are close to the ground truth (Figure 5.8). DPDNet makes large errors as it is trained on Canon data and does not generalize. We improve the accuracy of DPDNet by providing vignetting corrected images as input, but its accuracy is still lower than ours.

Defocus map estimation results. Our method produces defocus maps that are closest to the ground truth (Figure 5.9), especially on textureless regions, such as the toy and clock in the first scene. Similar to [144], depth accuracy near edges can be improved by



(a) input image (b) ground truth (c) ours (d) ours w/ GF (e) Wiener [194] (f) DMENet [100] (g) [144] (h) Garg [53] (i) Wadhwa [181]

Figure 5.9: **Qualitative comparisons of defocus map estimation methods.** Input images (a) shown as the average of two DP views, ground truth defocus maps (b) from focus stacks with zero confidence pixels in white, our defocus maps (c), and our defocus maps with guided filtering (d), and defocus maps from other methods (f)-(i). Overall, our method produces results that are closest to the ground truth, and correctly handles textureless regions as well.

method	all-in-focus image			defocus map		
	PSNR ↑	SSIM ↑	MAE ↓	AIWE(1) ↓	AIWE(2) ↓	$1-\lvert\rho_s\rvert$ ↓
Wiener deconv. [194]	25.806	0.704	0.032	0.156	0.197	0.665
DPDNet [1]	25.591	0.777	0.034	-	-	-
DMENet [100]	-	-	-	0.144	0.183	0.586
Punnappurath *et al.* [144]	-	-	-	0.124	0.161	0.444
Garg *et al.* [53]	-	-	-	0.079	0.102	0.208
Wadhwa *et al.* [181]	-	-	-	0.141	0.177	0.540
ours	26.692	0.804	0.027	0.047	0.076	0.178
ours w/ guided filtering	26.692	0.804	0.027	0.059	0.083	0.193

Table 5.1: **Quantitative evaluations of defocus deblurring and defocus map estimation methods on our DP dataset.** "-" indicates not applicable. We use the affine-invariant metrics from [53] for defocus map evaluation. Our method achieves the best performance (highlighted in red) in both tasks.

guided filtering [69] as shown in Figure 5.9(d).

Ablation studies. We investigate the effect of each loss function term by removing them one at a time. We show quantitative results in Table 5.2, and qualitative comparisons in Figure 5.10. Our full pipeline has the best overall performance in recovering all-in-focus images and defocus maps. $\mathcal{L}_{\text{intensity}}$ and $\mathcal{L}_{\text{alpha}}$ strongly affect the smoothness of all-in-

method	all-in-focus image			defocus map		
	PSNR ↑	SSIM ↑	MAE ↓	AIWE(1) ↓	AIWE(2) ↓	$1 - \lvert\rho_s\rvert$ ↓
full	26.692	0.804	0.027	0.047	0.076	0.178
no $\mathcal{L}_{\text{intensity}}$	14.882	0.158	0.136	0.047	0.078	0.185
no $\mathcal{L}_{\text{alpha}}$	24.748	0.726	0.037	0.161	0.206	0.795
no $\mathcal{L}_{\text{entropy}}$	27.154	0.819	0.026	0.057	0.085	0.190
no $\mathcal{L}_{\text{aux}}$	26.211	0.768	0.030	0.148	0.190	0.610
no $\mathcal{B}$	26.265	0.790	0.028	0.063	0.092	0.214

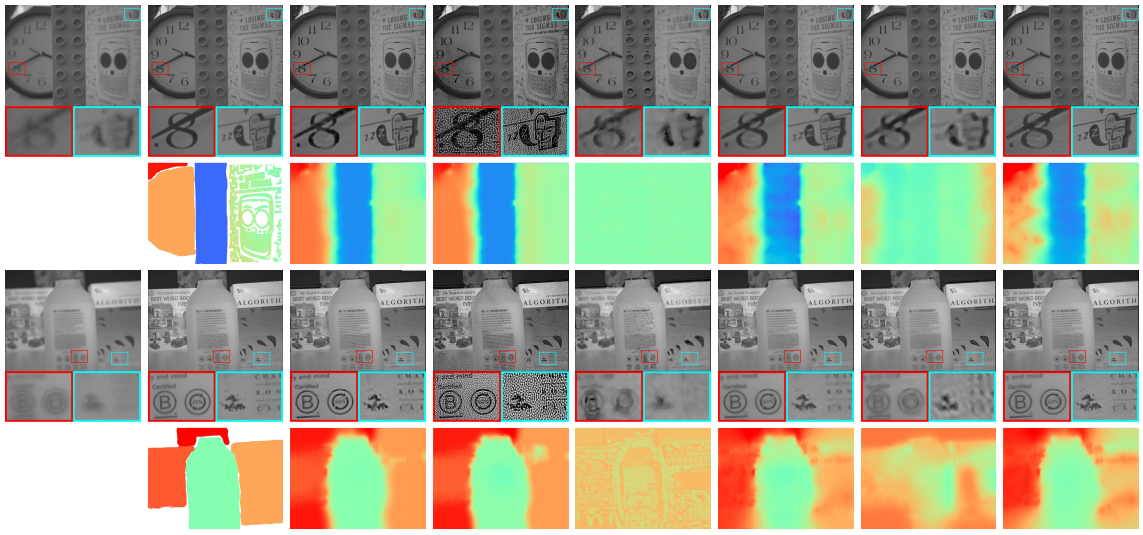
Table 5.2: **Quantitative comparisons of ablation studies.** We compare the full pipeline with removals of the regularization terms $\mathcal{L}_{\text{alpha}}$, $\mathcal{L}_{\text{intensity}}$ and $\mathcal{L}_{\text{entropy}}$, the auxiliary data loss $\mathcal{L}_{\text{aux}}$, and bias correction term $\mathcal{B}$ respectively. For all ablation experiments, we set the weights on remaining terms to be the same as the ones in the full pipeline. Best and second best results are highlighted in red and orange.

focus images and defocus maps, respectively. Without $\mathcal{L}_{\text{entropy}}$ or $\mathcal{L}_{\text{aux}}$, even though recovered all-in-focus images are reasonable, scene content is smeared across multiple MPI layers, leading to incorrect defocus maps. Finally, without the bias correction term $\mathcal{B}$, defocus maps are biased towards smaller blur radii, especially in textureless areas where noise is more apparent, e.g., the white clock area. We further show quantitatively that without $\mathcal{B}$, the estimated defocus size is smaller on average as predicted by our analysis (Figure 5.11).

Results on data from Abuolaim and Brown [1]. Even though Abuolaim and Brown [1] train their model on data from a Canon camera, they also capture Pixel 4 data for qualitative tests. We run our method on their Pixel 4 data, using the calibration from our device, and show that our recovered all-in-focus image has fewer artifacts (Figure 5.12). This demonstrates that our method generalizes well across devices of the same model, even without re-calibration.

5.5 Discussion and Summary

We presented a method that optimizes an MPI scene representation to jointly recover a defocus map and an all-in-focus image from a single dual-pixel capture. We showed



(a) input image (b) ground truth (c) ours full (d) no $\mathcal{L}_{\text{intensity}}$ (e) no $\mathcal{L}_{\text{alpha}}$ (f) no $\mathcal{L}_{\text{entropy}}$ (g) no $\mathcal{L}_{\text{aux}}$ (h) no $\mathcal{B}$

Figure 5.10: **Ablation studies.** Input images (a), ground truth all-in-focus images, and defocus maps (b) with zero confidence pixels in white, our results (c), and our results with different terms removed one at a time (d)-(h). Removing $\mathcal{L}_{\text{intensity}}$ and $\mathcal{L}_{\text{alpha}}$ strongly affects the smoothness of all-in-focus images and defocus maps respectively. Results without entropy regularization $\mathcal{L}_{\text{entropy}}$, $\mathcal{L}_{\text{aux}}$, or the bias correction $\mathcal{B}$, exhibit more errors in defocus maps on textureless regions (clock).

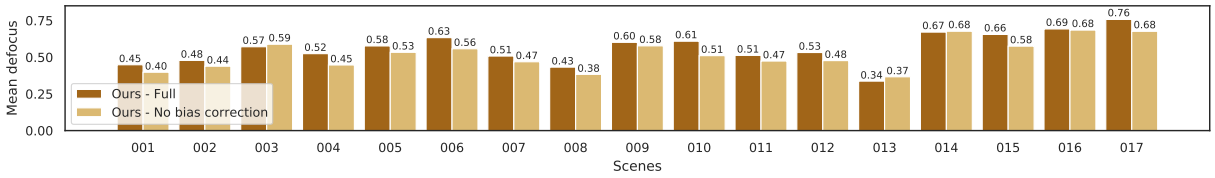
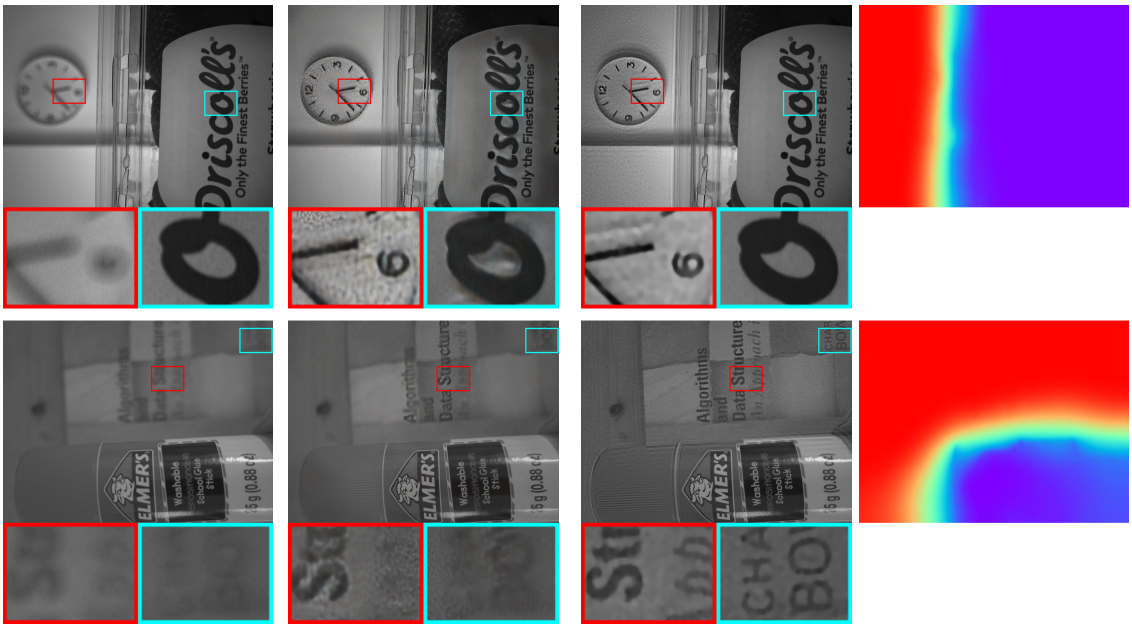


Figure 5.11: **Quantitative evaluation of the effect of the bias correction term $\mathcal{B}$ through an ablation study.** We show the mean of the predicted defocus map for our full pipeline versus an ablation where bias correction term is not applied. Defocus is measured as the relative scaling applied to the calibrated kernels. Without bias correction $\mathcal{B}$, the mean defocus is lower in 14 of the 17 scenes, i.e., the prediction is biased towards smaller defocus size.

that image noise introduces a bias in the optimization that, under suitable assumptions, can be quantified and corrected for. We also introduced additional priors to regular-



(a) input from [1] (b) DPDNet [1] (c) our results

Figure 5.12: **Generalization to other Google Pixel 4 cameras.** We show results on data from [1]. Our method recovers all-in-focus images with fewer artifacts, while using the calibration data from our device.

ize the optimization, and showed their effectiveness via ablation studies. Our method improves upon past work on both defocus map estimation and blur removal, when evaluated on a new dataset we captured with a consumer smartphone camera.

We discuss some limitations of our method, which suggest directions for future research. First, our method does not require a large dataset with ground truth to train on, but still relies on a one-time blur kernel calibration procedure. It would be interesting to explore blind deconvolution techniques [49, 103] that can simultaneously recover the all-in-focus image, defocus map, and unknown blur kernels, thus removing the need for kernel calibration. The development of parametric blur kernel models that can accurately reproduce the features we observed (i.e., spatial variation, lack of symmetry, lack of circularity) can facilitate this research direction. Second, the MPI representation discretizes the scene into a set of fronto-parallel depth layers. This can potentially result in discretization artifacts in scenes with continuous depth variation. In practice, we did not find this to be an issue, thanks to the use of the soft-blending operation to synthesize the all-in-focus image and defocus map. Nevertheless, it could be useful to

replace the MPI representation with a continuous one, e.g., neural radiance fields [121], to help better model continuously-varying depth. Third, reconstructing an accurate all-in-focus image becomes more difficult as defocus blur increases (e.g., very distant scenes at non-infinity focus) and more high-frequency content is missing from the input image. This is a fundamental limitation shared among all deconvolution techniques. Using powerful data-driven priors to hallucinate the missing high frequency content (e.g., deep-learning-based deconvolution techniques) can help alleviate this limitation. Fourth, the high computational complexity of our technique makes it impractical for real-time operation, especially on resource-constrained devices such as smartphones. Therefore, it is worth exploring optimized implementations.

Chapter 6

Conclusion and Future Directions

In this thesis, we apply differential imaging to solve three challenging 3D reconstruction problems: NLOS imaging, specular object reconstruction, and single-shot depth from defocus. We have shown that differential imaging enables local differential analysis through interpolation and imaging systems with small form-factors.

In Chapter 2, we introduce a novel theory of Fermat paths—light paths that satisfy Fermat's principle. Fermat paths encode rich geometric information of the shape of interest, and their pathlength function can be thought of as a generalized multi-branched SDF. Based on the properties of Fermat paths through first- and second-order differential analyses, we propose surface reconstruction algorithms based on measurements of lengths or directions of Fermat paths.

In Chapter 3, we propose a geometric solution to NLOS imaging based on the theory of Fermat paths. We apply differential imaging by densely scanning the visible surface using a transient imaging system and extract, from these transient measurements, a colletion of pathlength information of the Fermat paths travelled between the visible surface and the hidden object. We then used the Fermat flow algorithm developed in Chapter 2 to reconstruct accurate NLOS shapes.

In Chapter 4, we present a method for recovering the shape of a specular mirror-like surface that is illuminated by a near-field point light source and is imaged by a differentially moving camera. The theory of Fermat paths allows us to reconstruct a OPF of surfaces. We then propose a method to resolve the remaining ambiguity using radiometric information of the moving specularities.

In Chapter 5, we propose a method to simultaneously estimate the defocus map and recover the latent all-in-focus image from a single dual-pixel (DP) image. The differen-

tial baseline in the left and right views of a DP image provides more constaints than a single capture using a conventional sensor. Therefore, the single-shot depth from defocus problem becomes less ill-posed and our optimization method achieves state-of-the-art performance by inverting the image formation with the help of a few regularization terms.

6.1 Future Directions

Fermat paths for complex scene reconstruction. In both NLOS imaging (Chapter 3) and specular object reconstruction (Chapter 4), we have presented results on only simple smooth surfaces. Real-world scenes are typically much more complex than single isolated objects. Higher scene complexity leads to the following two issues when applying the theory of Fermat paths.

First, extracting information of Fermat paths become challenging. Simple smooth objects generate transient measurements with clearly detectable discontinuities or isolated specularities. As scene complexity increases, there will likely be more transient discountinuties or specularities present in the measurement and even affecting each other, e.g., two transient discontinuities or specularities merge into one and violating the uniqueness constraint of the Fermat flow algorithm.

Second, differential analysis should be applied separately to each branch of the multi-valued Fermat paths function. A simple scene tends to create less branches that are well-separable, whereas a complex scene may lead to overlapping branches—making the separation extremely challenging. One needs to properly establish correspondences across measurements at all scan points before applying the Fermat flow reconstruction algorithm.

Handling the above issues will allow the reconstruction of more complex scenes in both NLOS imaging and specular object reconstruction. One possible solution is to use the reconstruction results from other methods as guidance and refine their results using the theory of Fermat paths. In the case of NLOS imaging, one could consider combining our method with volumetric methods that use all transient measurements to reconstruct an albedo volume by solving an inverse radiometric problem [5, 113, 136, 179]. Volumetric methods usually produce lower resolution results than geometric methods like ours, but are better at handling complex scenes. Therefore, one might

consider using their reconstructions to facilitate the detection and branch assignement of Fermat paths, and then use the Fermat flow algorithm to refine the geometry.

Fermat paths for transparent object reconstruction. For transparent objects, the refraction of light follows Snell's law, which is also derived from Fermat's principle. As a result, all of our theory and analyses of Fermat paths of single-bounce reflection naturally apply to single-bounce refraction as well. One needs to account for refractive indices when generalizing the theory to refraction events. A potential application is the reconstruction of dynamic fluid surfaces by placing the light source or the sensor under water [123] and diffrentially translating one of the two.

Experimental validation of the radiometric reconstruction method for specular object reconstruction. In Chapter 4, we propose to use radiometric information of specularites to disambiguate within the OPF of specular surfaces reconstructed using the integral form of the Fermat flow algorithm. We conduct theoretical analysis on reducing the ambiguity using a pair/collection of image irradiances of specularities.

A natural next step is to build a hardware prototype and conduct real experiments to fully verify the effectiveness of the proposed method. In fact, we once built a hardware implmentation, but noticed an unneglectable deviation of the measured irradiances of specularities from the theoretical model (although the model has been verified through a physics-based renderer), which caused trouble to disambiguating within the of surfaces. We speculate that the deviation was due to many factors not aligning with the assumptions of the imaging setting. First, the light source was not a perfect point light source and had a non-uniform intensity profile. Second, we noticed that radiometric measurements are very sensitive to any surface imperfection such as scratches and dents on the specular surface. We envision as a future work to carefully build a prototype and robustly handle these real-world imaging challenges, which, if sucessful, could open the door to a practical shape from shading technique tailored for mirror surfaces.

Combining differential imaging and other imaging modalities. It is worth exploring how we may combine differential imaging with other imaging modalities. For example, smartphones, such as iPhone 14 Pro, now adopts quad-pixel (QP) sensors, which can be seen as four-sample light field cameras. Although these RGB QP images do not provide depth information directly, we may apply differential imaging to infer depth, similar to how autofocus works and how we conduct depth from defocus using DP images. And such a depth map would be of high spatial resolution, but with relatively low

certainty. On the other hand, the Lidar scanner on iPhone provides more accurate depth information, but is sparse and can suffer from noise of time-of-flight sensors. If we develop an algorithm that combines the depth cues provided by both differential RGB images and the Lidar data, then smartphones may have a more refined depth estimation pipeline for downstream tasks, leading to new capabilities of 3D scene understanding on mobile devices.

Appendix A

Appendix to Chapter 2

In this appendix, we cover the following topics:

1. In Section A.1, we prove Proposition 6 (Fermat's principle).
2. In Section A.2, we prove Proposition 8 (Fermat flow equation).
3. In Section A.3, we prove Proposition 10.
4. In Section A.4, we discuss how to estimate the gradients used in the Fermat flow equation in the case of a non-planar manifold $\mathcal{V}$.
5. In Section A.5, we discuss our approach for surface fitting under specular path-length constraints.
6. In Section A.6, we prove Proposition 6′, which is the the analogue of Proposition 6 for the non-confocal case.
7. In Section A.7, we prove Proposition 8′, which is the the analogue of Proposition 8 for the non-confocal case.
8. In Section A.8, we prove Proposition 10′, which is the the analogue of Proposition 10 for the non-confocal case.
9. In Section A.9, we derive the incident and outgoing pathlengths of a Fermat path given its total pathlength.

A.1 Proof of Proposition 6

We prove Proposition 6, which we restate here for convenience.

Proposition 6. Fermat's principle. *Let* $(p,q) \in [0,1]^2$ *be a parameterization of the surface* $\mathcal{X}$*. Then, for any point* $\boldsymbol{v}$*,*

$$\mathcal{S}(\boldsymbol{v}) = \left\{ \boldsymbol{x} \in \mathcal{X} : \nabla_{(p,q)} \tau(\boldsymbol{x}(p,q); \boldsymbol{v}) = \boldsymbol{0} \right\}. \tag{A.1}$$

Let $r \in [0,1]$ *be a parameterization of the surface boundary* $\partial\mathcal{X}$*. Then, for any point* $\boldsymbol{v}$*,*

$$\mathcal{B}(\boldsymbol{v}) = \{ \boldsymbol{x} \in \partial\mathcal{X} : \partial\tau(\boldsymbol{x}(r); \boldsymbol{v}) / \partial r = 0 \}. \tag{A.2}$$

Proof. For the first part of the proposition, Equation (A.1), we have,

$$\frac{\partial \tau(\boldsymbol{x}(p,q); \boldsymbol{v})}{\partial p} = \left\langle \frac{\boldsymbol{x}(p,q) - \boldsymbol{v}}{\|\boldsymbol{x}(p,q) - \boldsymbol{v}\|}, \boldsymbol{x}_p(p,q) \right\rangle = \frac{4}{\tau(\boldsymbol{x}(p,q); \boldsymbol{v})} \langle \boldsymbol{x}(p,q) - \boldsymbol{v}, \boldsymbol{x}_p(p,q) \rangle \tag{A.3}$$

where $\boldsymbol{x}_p(p,q)$ is the partial derivative of $\boldsymbol{x}$ with respect to p. The vector $\boldsymbol{x}_p(p,q)$ is tangent to the surface $\mathcal{X}$ at $\boldsymbol{x}$, and therefore orthogonal to the surface normal $\hat{\boldsymbol{n}}(\boldsymbol{x})$ at that point. If $\boldsymbol{x}(p,q) \in \mathcal{S}(\boldsymbol{v})$, then from the specular reflection property, the vector $\boldsymbol{x}(p,q) - \boldsymbol{v}$ is parallel to the normal $\hat{\boldsymbol{n}}(\boldsymbol{x})$. Therefore,

$$\boldsymbol{x}(p,q) - \boldsymbol{v} \parallel \hat{\boldsymbol{n}}(\boldsymbol{x}) \text{ and } \boldsymbol{x}_p(p,q) \perp \hat{\boldsymbol{n}}(\boldsymbol{x}) \Rightarrow \langle \boldsymbol{x}(p,q) - \boldsymbol{v}, \boldsymbol{x}_p(p,q) \rangle = 0, \tag{A.4}$$

and from Equation (A.3),

$$\frac{\partial \tau(\boldsymbol{x}(p,q); \boldsymbol{v})}{\partial p} = 0. \tag{A.5}$$

The proof for $\frac{\partial \tau(\boldsymbol{x}(p,q); \boldsymbol{v})}{\partial q} = 0$ is exactly the same. Therefore, if $\boldsymbol{x}(p,q) \in \mathcal{S}(\boldsymbol{v})$, then $\nabla_{(p,q)} \tau(\boldsymbol{x}(p,q); \boldsymbol{v}) = \boldsymbol{0}$. Conversely, if $\nabla_{(p,q)} \tau(\boldsymbol{x}(p,q); \boldsymbol{v}) = \boldsymbol{0}$, from Equation (A.3) we have that either $\boldsymbol{v} = \boldsymbol{x}(p,q)$, or that $\boldsymbol{x}(p,q) - \boldsymbol{v}$ is orthogonal to both $\boldsymbol{x}_p(p,q)$ and $\boldsymbol{x}_q(p,q)$, and therefore parallel to the normal $\hat{\boldsymbol{n}}(\boldsymbol{x})$. By assuming that $\boldsymbol{v} \notin \mathcal{X}$, this implies that $\boldsymbol{x}(p,q) \in \mathcal{S}(\boldsymbol{v})$. This concludes the proof of the first part of the proposition.

For the second part of the proposition, Equation (A.2), we have,

$$\frac{\partial \tau(\boldsymbol{x}(r); \boldsymbol{v})}{\partial r} = \left\langle \frac{\boldsymbol{x}(r) - \boldsymbol{v}}{\|\boldsymbol{x}(r) - \boldsymbol{v}\|}, \boldsymbol{x}_r(r) \right\rangle = \frac{4}{\tau(\boldsymbol{x}(r); \boldsymbol{v})} \langle \boldsymbol{x}(r) - \boldsymbol{v}, \boldsymbol{x}_r(r) \rangle \tag{A.6}$$

where $\boldsymbol{x}_r(r)$ is the partial derivative of $\boldsymbol{x}$ with respect to r. The vector $\boldsymbol{x}_r(r)$ is parallel to the tangent $\hat{\boldsymbol{t}}(\boldsymbol{x})$ of the curve $\partial\mathcal{X}$ at $\boldsymbol{x}$. If $\boldsymbol{x}(r) \in \mathcal{B}(\boldsymbol{v})$, then from the property of boundary Fermat paths, the vector $\boldsymbol{x}(r) - \boldsymbol{v}$ is orthogonal to the tangent $\hat{\boldsymbol{t}}(\boldsymbol{x})$. Therefore,

$$\boldsymbol{x}(r) - \boldsymbol{v} \perp \hat{\boldsymbol{t}}(\boldsymbol{x}) \text{ and } \boldsymbol{x}_r(r) \parallel \hat{\boldsymbol{t}}(\boldsymbol{x}) \Rightarrow \langle \boldsymbol{x}(r) - \boldsymbol{v}, \boldsymbol{x}_r(r) \rangle = 0, \tag{A.7}$$

and from Equation (A.6),

$$\frac{\partial \tau\left(\boldsymbol{x}\left(r\right);\boldsymbol{v}\right)}{\partial r}=0. \tag{A.8}$$

Conversely, if $\frac{\partial \tau(\boldsymbol{x}(r);\boldsymbol{v})}{\partial r}=0$, from Equation (A.3) we have that either $\boldsymbol{v}=\boldsymbol{x}\left(r\right)$, or that $\boldsymbol{x}\left(r\right)-\boldsymbol{v}$ is orthogonal to $\boldsymbol{x}_r\left(r\right)$, and therefore orthogonal to $\hat{\boldsymbol{t}}\left(\boldsymbol{x}\right)$. By assuming that $\boldsymbol{v}\notin\partial\mathcal{X}$, this implies that $\boldsymbol{x}\left(r\right)\in\mathcal{B}\left(\boldsymbol{v}\right)$. This concludes the proof of the second part of the proposition. □

A.2 Proof of Proposition 8—the Fermat flow equation

Proposition 8. The Fermat flow equation. *Consider a branch of the Fermat pathlength function $\tau_{\mathcal{F}}\left(\boldsymbol{v}\right)$ evaluated at $\boldsymbol{v}$. Assume that there is a unique point $\boldsymbol{x}_{\mathcal{F}}\in\mathcal{F}\left(\boldsymbol{v}\right)$ with $\tau\left(\boldsymbol{x}_{\mathcal{F}};\boldsymbol{v}\right)=\tau_{\mathcal{F}}\left(\boldsymbol{v}\right)$. Then,*

$$\nabla_{\boldsymbol{v}}\tau_{\mathcal{F}}\left(\boldsymbol{v}\right)=-2\frac{\boldsymbol{x}_{\mathcal{F}}-\boldsymbol{v}}{\left\|\boldsymbol{x}_{\mathcal{F}}-\boldsymbol{v}\right\|}. \tag{A.9}$$

Proof. We will be using $\boldsymbol{v}=\left[v^x,v^y,v^z\right]^T$ to denote the 3D coordinates of the point $\boldsymbol{v}$, and similarly for all other vectors.

We first prove the proposition for the case of a *specular* path, that is, $\boldsymbol{x}_{\mathcal{F}}\in\mathcal{S}\left(\boldsymbol{v}\right)$. Let $\left(p,q\right)\in\left[0,1\right]^2$ be a parameterization of the surface $\mathcal{X}$ in a neighborhood around $\boldsymbol{x}_{\mathcal{F}}$, such that $\boldsymbol{x}_{\mathcal{F}}=\boldsymbol{x}\left(p\left(\boldsymbol{v}\right),q\left(\boldsymbol{v}\right)\right)$. We consider each coordinate of the vector $\nabla_{\boldsymbol{v}}\tau_{\mathcal{F}}\left(\boldsymbol{v}\right)$ separately. For the first coordinate, we have

$$\frac{\partial\tau_{\mathcal{F}}\left(\boldsymbol{v}\right)}{\partial v^x}=2\frac{\partial\left\|\boldsymbol{x}_{\mathcal{F}}-\boldsymbol{v}\right\|}{\partial v^x} \tag{A.10}$$

$$=2\frac{\partial\left\|\boldsymbol{x}\left(p\left(\boldsymbol{v}\right),q\left(\boldsymbol{v}\right)\right)-\boldsymbol{v}\right\|}{\partial v^x} \tag{A.11}$$

$$=2\left\langle\frac{\boldsymbol{x}\left(p\left(\boldsymbol{v}\right),q\left(\boldsymbol{v}\right)\right)-\boldsymbol{v}}{\left\|\boldsymbol{x}\left(p\left(\boldsymbol{v}\right),q\left(\boldsymbol{v}\right)\right)-\boldsymbol{v}\right\|},\frac{\partial\left(\boldsymbol{x}\left(p\left(\boldsymbol{v}\right),q\left(\boldsymbol{v}\right)\right)-\boldsymbol{v}\right)}{\partial v^x}\right\rangle \tag{A.12}$$

$$=2\left\langle\frac{\boldsymbol{x}_{\mathcal{F}}-\boldsymbol{v}}{\left\|\boldsymbol{x}_{\mathcal{F}}-\boldsymbol{v}\right\|},\boldsymbol{x}_p\left(p\left(\boldsymbol{v}\right),q\left(\boldsymbol{v}\right)\right)\frac{\partial p\left(\boldsymbol{v}\right)}{\partial v^x}+\boldsymbol{x}_q\left(p\left(\boldsymbol{v}\right),q\left(\boldsymbol{v}\right)\right)\frac{\partial q\left(\boldsymbol{v}\right)}{\partial v^x}-\left[1,0,0\right]^T\right\rangle \tag{A.13}$$

$$=\frac{2}{\left\|\boldsymbol{x}_{\mathcal{F}}-\boldsymbol{v}\right\|}\left(\left\langle\boldsymbol{x}_{\mathcal{F}}-\boldsymbol{v},\boldsymbol{x}_p\left(p\left(\boldsymbol{v}\right),q\left(\boldsymbol{v}\right)\right)\right\rangle\frac{\partial p\left(\boldsymbol{v}\right)}{\partial v^x}+\left\langle\boldsymbol{x}_{\mathcal{F}}-\boldsymbol{v},\boldsymbol{x}_q\left(p\left(\boldsymbol{v}\right),q\left(\boldsymbol{v}\right)\right)\right\rangle\frac{\partial q\left(\boldsymbol{v}\right)}{\partial v^x}-\left\langle\boldsymbol{x}_{\mathcal{F}}-\boldsymbol{v},\left[1,0,0\right]\right.\right. \tag{A.14}$$

In Equation (A.14), the vectors $\boldsymbol{x}_p\left(p\left(\boldsymbol{v}\right),q\left(\boldsymbol{v}\right)\right)$ and $\boldsymbol{x}_q\left(p\left(\boldsymbol{v}\right),q\left(\boldsymbol{v}\right)\right)$ are tangent to the surface $\mathcal{X}$ at $\boldsymbol{x}_{\mathcal{F}}$. Given that we assumed that $\boldsymbol{x}_{\mathcal{F}}\in\mathcal{S}\left(\boldsymbol{v}\right)$, and from the definition of the specular set, the vector $\boldsymbol{x}_{\mathcal{F}}-\boldsymbol{v}$ is parallel to the normal $\hat{\boldsymbol{n}}\left(\boldsymbol{x}_{\mathcal{F}}\right)$ of $\mathcal{X}$ at $\boldsymbol{x}_{\mathcal{F}}$. Therefore, $\boldsymbol{x}_{\mathcal{F}}-\boldsymbol{v}$ is orthogonal to $\boldsymbol{x}_p\left(p\left(\boldsymbol{v}\right),q\left(\boldsymbol{v}\right)\right)$ and $\boldsymbol{x}_q\left(p\left(\boldsymbol{v}\right),q\left(\boldsymbol{v}\right)\right)$, and Equation (A.14)

becomes

$$\frac{\partial \tau_{\mathcal{F}}(\boldsymbol{v})}{\partial v^x} = \frac{2}{\|\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}\|}\left(0 \cdot \frac{\partial p(\boldsymbol{v})}{\partial v^x} + 0 \cdot \frac{\partial q(\boldsymbol{v})}{\partial v^x} - \left\langle \boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}, [1,0,0]^T \right\rangle\right) = -2\frac{(\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v})^x}{\|\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}\|}. \tag{A.15}$$

Exactly analogously, we can prove that

$$\frac{\partial \tau_{\mathcal{F}}(\boldsymbol{v})}{\partial v^y} = -2\frac{(\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v})^y}{\|\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}\|} \qquad \text{and} \qquad \frac{\partial \tau_{\mathcal{F}}(\boldsymbol{v})}{\partial v^z} = -2\frac{(\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v})^z}{\|\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}\|}. \tag{A.16}$$

Combining Equations (A.15) and (A.16) completes the proof for the specular case.

We now prove the proposition for the case of a *boundary* path, that is, $\boldsymbol{x}_{\mathcal{F}} \in \mathcal{B}(\boldsymbol{v})$. Let $r \in [0,1]$ be a parameterization of the surface boundary $\partial\mathcal{X}$ in a neighborhood around $\boldsymbol{x}_{\mathcal{F}}$, such that $\boldsymbol{x}_{\mathcal{F}} = \boldsymbol{x}(r(\boldsymbol{v}))$. We again consider each coordinate of the vector $\nabla_v \tau_{\mathcal{F}}(\boldsymbol{v})$ separately. For the first coordinate, we have

$$\begin{aligned}
\frac{\partial \tau_{\mathcal{F}}(\boldsymbol{v})}{\partial v^x} &= 2\frac{\partial \|\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}\|}{\partial v^x} && \text{(A.17)}\\
&= 2\frac{\partial \|\boldsymbol{x}(r(\boldsymbol{v})) - \boldsymbol{v}\|}{\partial v^x} && \text{(A.18)}\\
&= 2\left\langle \frac{\boldsymbol{x}(r(\boldsymbol{v})) - \boldsymbol{v}}{\|\boldsymbol{x}(r(\boldsymbol{v})) - \boldsymbol{v}\|}, \frac{\partial(\boldsymbol{x}(r(\boldsymbol{v})) - \boldsymbol{v})}{\partial v^x} \right\rangle && \text{(A.19)}\\
&= 2\left\langle \frac{\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}}{\|\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}\|}, \boldsymbol{x}_r(r(\boldsymbol{v}))\frac{\partial r(\boldsymbol{v})}{\partial v^x} - [1,0,0]^T \right\rangle && \text{(A.20)}\\
&= \frac{2}{\|\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}\|}\left(\langle \boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}, \boldsymbol{x}_r(r(\boldsymbol{v}))\rangle \frac{\partial r(\boldsymbol{v})}{\partial v^x} - \left\langle \boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}, [1,0,0]^T \right\rangle\right). && \text{(A.21)}
\end{aligned}$$

In Equation (A.21), the vector $\boldsymbol{x}_r(r(\boldsymbol{v}))$ is parallel to the tangent $\hat{\boldsymbol{t}}(\boldsymbol{x}_{\mathcal{F}})$ of the surface boundary $\partial\mathcal{X}$ at $\boldsymbol{x}_{\mathcal{F}}$. Given that we assumed that $\boldsymbol{x}_{\mathcal{F}} \in \mathcal{B}(\boldsymbol{v})$, and from the definition of the boundary set, the vector $\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}$ is orthogonal to the tangent $\hat{\boldsymbol{t}}(\boldsymbol{x}_{\mathcal{F}})$. Therefore, $\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}$ is also orthogonal to $\boldsymbol{x}_r(r(\boldsymbol{v}))$, and Equation (A.21) becomes

$$\frac{\partial \tau_{\mathcal{F}}(\boldsymbol{v})}{\partial v^x} = \frac{2}{\|\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}\|}\left(0 \cdot \frac{\partial r(\boldsymbol{v})}{\partial v^x} - \left\langle \boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}, [1,0,0]^T \right\rangle\right) = -2\frac{(\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v})^x}{\|\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}\|}. \tag{A.22}$$

Exactly analogously, we can prove that

$$\frac{\partial \tau_{\mathcal{F}}(\boldsymbol{v})}{\partial v^y} = -2\frac{(\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v})^y}{\|\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}\|} \qquad \text{and} \qquad \frac{\partial \tau_{\mathcal{F}}(\boldsymbol{v})}{\partial v^z} = -2\frac{(\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v})^z}{\|\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}\|}. \tag{A.23}$$

Combining Equations (A.22) and (A.23) completes the proof for the boundary case. □

A.3 Proof of Proposition 10

Proposition 10. *Let a point $\boldsymbol{x}_{\mathcal{S}} \in \mathcal{S}(\boldsymbol{v})$ belong to the specular set. If $\kappa_{min}, \kappa_{max}$ are the principal curvatures of $\mathcal{X}$ at $\boldsymbol{x}_{\mathcal{S}}$, then:*

- *If $\tau_{\mathcal{S}}$ is a local minimum of $\tau(\boldsymbol{x}; \boldsymbol{v})$, $2/\tau_{\mathcal{S}} < \kappa_{min}$.*
- *If $\tau_{\mathcal{S}}$ is a local maximum of $\tau(\boldsymbol{x}; \boldsymbol{v})$, $\kappa_{max} < 2/\tau_{\mathcal{S}}$.*
- *If $\tau_{\mathcal{S}}$ is a saddle point of $\tau(\boldsymbol{x}; \boldsymbol{v})$, $\kappa_{min} \leq 2/\tau_{\mathcal{S}} \leq \kappa_{max}$.*

Proof. As $\tau_{\mathcal{S}}$ is a discontinuity corresponding to a specular path, the sphere $\text{Sph}(\rho_{\mathcal{S}}; \boldsymbol{v})$ of radius $\rho_{\mathcal{S}} = \tau_{\mathcal{S}}/2$ and center $\boldsymbol{v}$ will be tangent to the surface at point $\boldsymbol{x}_{\mathcal{S}}$.

We consider first the case where $\tau_{\mathcal{S}}$ is a local minimum of $\tau(\boldsymbol{x}; \boldsymbol{v})$. Then, there will be some neighborhood $\mathcal{N}(\boldsymbol{x}_{\mathcal{S}}) \subset \mathcal{X}$ of $\boldsymbol{x}_{\mathcal{S}}$ on $\mathcal{X}$ such that, for all points $\boldsymbol{x} \in \mathcal{N}(\boldsymbol{x}_{\mathcal{S}})$, $\tau(\boldsymbol{x}; \boldsymbol{v}) \geq \tau_{\mathcal{S}}$. Equivalently, all of $\mathcal{N}(\boldsymbol{x}_{\mathcal{S}})$ lies outside the sphere $\text{Sph}(\rho_{\mathcal{S}}; \boldsymbol{v})$, and is tangent to that sphere at the point $\boldsymbol{x}_{\mathcal{S}}$. Therefore, the curves on $\mathcal{N}(\boldsymbol{x}_{\mathcal{S}})$ passing through $\boldsymbol{x}_{\mathcal{S}}$ in all possible tangent directions are also tangent to and outside of the sphere $\text{Sph}(\rho_{\mathcal{S}}; \boldsymbol{v})$. Consequently, all normal curvatures of $\mathcal{X}$ at $\boldsymbol{x}_{\mathcal{S}}$ are greater than the inverse of the radius of $\text{Sph}(\rho_{\mathcal{S}}; \boldsymbol{v})$. From the definition of principal curvatures, we conclude that $2/\tau_{\mathcal{S}} = \rho_{\mathcal{S}} \leq \kappa_{\text{min}}$.

The case when $\tau_{\mathcal{S}}$ is a local maximum of $\tau(\boldsymbol{x}; \boldsymbol{v})$ proceeds very similarly: In this case, all of $\mathcal{N}(\boldsymbol{x}_{\mathcal{S}})$ is *inside* the sphere $\text{Sph}(\rho_{\mathcal{S}}; \boldsymbol{v})$, and tangent to it at $\boldsymbol{x}_{\mathcal{S}}$. Therefore, the curves on $\mathcal{N}(\boldsymbol{x}_{\mathcal{S}})$ passing through $\boldsymbol{x}_{\mathcal{S}}$ in all possible tangent directions are also tangent to and *inside* of the sphere $\text{Sph}(\rho_{\mathcal{S}}; \boldsymbol{v})$. Consequently, all normal curvatures of $\mathcal{X}$ at $\boldsymbol{x}_{\mathcal{S}}$ are *smaller* than the inverse of the radius of $\text{Sph}(\rho_{\mathcal{S}}; \boldsymbol{v})$. From the definition of principal curvatures, we conclude that $\kappa_{\text{max}} \leq \rho_{\mathcal{S}} = 2/\tau_{\mathcal{S}}$.

Finally, we proceed similarly for the case $\tau_{\mathcal{S}}$ is a saddle point of $\tau(\boldsymbol{x}; \boldsymbol{v})$: In this case, $\mathcal{N}(\boldsymbol{x}_{\mathcal{S}})$ lies partially inside and partially outside the sphere $\text{Sph}(\rho_{\mathcal{S}}; \boldsymbol{v})$. Therefore, while all the curves on $\mathcal{N}(\boldsymbol{x}_{\mathcal{S}})$ passing through $\boldsymbol{x}_{\mathcal{S}}$ in all possible tangent directions are still tangent to the sphere $\text{Sph}(\rho_{\mathcal{S}}; \boldsymbol{v})$; some of them will be inside and some outside of the $\text{Sph}(\rho_{\mathcal{S}}; \boldsymbol{v})$. Consequently, there exist normal curvatures of $\mathcal{X}$ at $\boldsymbol{x}_{\mathcal{S}}$ that are smaller, and others that are greater than the inverse of the radius of $\text{Sph}(\rho_{\mathcal{S}}; \boldsymbol{v})$. From the definition of principal curvatures, we conclude that $\kappa_{\text{min}} \leq \rho_{\mathcal{S}} = 2/\tau_{\mathcal{S}} \leq \kappa_{\text{max}}$. □

A.4 Gradient Estimation on General Scanning Manifolds

We can still estimate the gradient from perturbations of v on any general manifolds $\mathcal{V}$. For this, we assume that $\mathcal{V}$ can be parameterized by $(m,n) \in [0,1]^2$, and therefore $v = [v^x(m,n), v^y(m,n), v^z(m,n)]^T$. If $x_\mathcal{F}$ is specular, $x_\mathcal{F} \in \mathcal{S}(v)$ then, by performing a calculation similar to the one in Equations (A.10)-(A.15), we have

$$\frac{\partial \tau_\mathcal{F}(v)}{\partial m} = 2\frac{\partial \|x_\mathcal{F} - v\|}{\partial m} \tag{A.24}$$

$$= 2\frac{\partial \|x(p(v), q(v)) - v\|}{\partial m} \tag{A.25}$$

$$= 2\left\langle \frac{x(p(v), q(v)) - v}{\|x(p(v), q(v)) - v\|}, \frac{\partial (x(p(v), q(v)) - v)}{\partial m} \right\rangle \tag{A.26}$$

$$= 2\left\langle \frac{x_\mathcal{F} - v}{\|x_\mathcal{F} - v\|}, x_p(p(v), q(v))\frac{\partial p(v)}{\partial m} + x_q(p(v), q(v))\frac{\partial q(v)}{\partial v^x} - v_m \right\rangle \tag{A.27}$$

$$= \frac{2}{\|x_\mathcal{F} - v\|}\left(\langle x_\mathcal{F} - v, x_p(p(v), q(v))\rangle \frac{\partial p(v)}{\partial m} + \langle x_\mathcal{F} - v, x_q(p(v), q(v))\rangle \frac{\partial q(v)}{\partial m} - \langle x_\mathcal{F} - v \tag{A.28}$$

where $v_m = \left[\frac{\partial v^x(m,n)}{\partial m}, \frac{\partial v^y(m,n)}{\partial m}, \frac{\partial v^z(m,n)}{\partial m}\right]^T$ is tangent to the manifold $\mathcal{V}$ at v. Using the fact that the vectors $x_p(p(v), q(v))$ and $x_q(p(v), q(v))$ are tangent to the surface $\mathcal{X}$ at $x_\mathcal{F}$, and therefore orthogonal to $x_\mathcal{F} - v$, Equation (A.28) becomes

$$\frac{\partial \tau_\mathcal{F}(v)}{\partial m} = -2\left\langle \frac{x_\mathcal{F} - v}{\|x_\mathcal{F} - v\|}, v_m \right\rangle. \tag{A.29}$$

Exactly analogously, we have,

$$\frac{\partial \tau_\mathcal{F}(v)}{\partial n} = -2\left\langle \frac{x_\mathcal{F} - v}{\|x_\mathcal{F} - v\|}, v_n \right\rangle. \tag{A.30}$$

We can also prove Equations (A.29) and (A.30) for the boundary case, $x_\mathcal{F} \in \mathcal{B}(v)$, by likewise adapting the calculation of Equations (A.17)-(A.22).

We can select the parameterization (m,n) such that the tangent vectors v_m, v_n are of unit norm and orthogonal to each other (e.g., by making the tangent vectors be the principal curvature directions). Then Equations (A.29) and (A.30) correspond simply to rotating coordinate system axes, from the global coordinate system to the local coordinate system at v with axes $v_m, v_n, v_m \times v_n$. In this coordinate system, the gradient will

equal:

$$\left(\frac{\partial \tau_{\mathcal{F}}(\boldsymbol{v})}{\partial m}, \frac{\partial \tau_{\mathcal{F}}(\boldsymbol{v})}{\partial n}, \sqrt{4 - \left(\frac{\partial \tau_{\mathcal{F}}(\boldsymbol{v})}{\partial m} \right)^2 - \left(\frac{\partial \tau_{\mathcal{F}}(\boldsymbol{v})}{\partial n} \right)^2} \right), \tag{A.31}$$

where the first two terms can be estimated by perturbing the point $\boldsymbol{v}$ on the scanning manifold $\mathcal{V}$ and interpolating. Finally, by performing a rotation, we can obtain from Equation (A.31) the gradient $\nabla_{\boldsymbol{v}} \tau_{\mathcal{F}}(\boldsymbol{v})$ in the global coordinate system. For a planar manifold $\mathcal{V}$, the above discussion reduces to Equation (2.35).

A.5 Surface Fitting Under Specular Pathlength Constraints

We provide details about the surface fitting procedure discussed in Section 2.4.1. The fact that our reconstruction relies on interpolated estimates, rather than direct measurements, of $\nabla_{\boldsymbol{v}} \tau_{\mathcal{F}}(\boldsymbol{v})$ can potentially introduce some error in the reconstructed point cloud and fitted surface. We can improve the accuracy of an initial surface reconstruction by modifying it so that it more closely matches the set $\{\tau_{\mathcal{F}}(\boldsymbol{v}_m), m = 1, \ldots, M\}$ of Fermat pathlength measurements available to us. This requires solving an optimization problem of the form:

$$\min_{\mathcal{S}} \sum_{m=1}^{M} \left\| \tau_{\mathcal{F}}(\boldsymbol{v}_m) - \tau_{\mathcal{F}}(\mathcal{S}, \boldsymbol{v}_m) \right\|^2, \tag{A.32}$$

where for a surface $\mathcal{S}$ and a point $\boldsymbol{v}$, the function $\tau_{\mathcal{F}}(\mathcal{S}, \boldsymbol{v})$ returns the length of the Fermat path between $\boldsymbol{v}$ and points on $\mathcal{S}$. This is a challenging *geometric* optimization problem, because of the difficulty of evaluating $\tau_{\mathcal{F}}(\mathcal{S}, \boldsymbol{v})$, and because the function $\tau_{\mathcal{F}}(\mathcal{S}, \boldsymbol{v})$ can be multi-valued.

To simplify exposition, we consider the case where we perform fitting for a single Fermat pathlength measurement $\tau_{\mathcal{F}}(\boldsymbol{v})$. We also assume that the surface $\mathcal{S}$ is a triangular mesh of fixed topology, consisting of K triangles, $\mathcal{S} = \bigcup_{k=1}^{K} \mathcal{T}_k$, and V vertices. For each triangle $\mathcal{T}_k$, the 3×3 matrix $\boldsymbol{V}_k$ contains the 3D coordinates of the triangle's three vertices. We use $\boldsymbol{V}$ to denote the $3 \times V$ matrix containing the 3D coordinates of all mesh vertices.

Then, Equation (A.32) is simplified to:

$$\min_{\boldsymbol{V}} \left\| \tau_{\mathcal{F}}(\boldsymbol{v}) - \tau_{\mathcal{F}}(\mathcal{S}, \boldsymbol{v}) \right\|^2, \tag{A.33}$$

To *approximately* solve the optimization problem of Equation (A.33), we begin by

identifying specular points on the current mesh $\mathcal{S}$ with respect to $\boldsymbol{v}$. This involves solving the specular *forward projection* problem [3, 4, 12]):

$$\min_{\mathcal{T}_k, k=1,\ldots,K} \min_{\boldsymbol{x}\in\mathcal{T}_k} \left\| 1 - \left\langle \frac{\boldsymbol{v}-\boldsymbol{x}}{\|\boldsymbol{v}-\boldsymbol{x}\|}, \hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \right\rangle \right\|^2. \tag{A.34}$$

Within each triangle $\mathcal{T}_k$, the inner optimization problem of Equation (A.34) can be solved by parameterizing points on the triangle using barycentric coordinates and then using gradient descent on this parametric representation. To accelerate both the inner and outer optimization problems, we use the sub-triangle search and triangle-prunning techniques from Walter et al. [182]. We denote by $\mathcal{T}^*$ the triangle containing the point $\boldsymbol{x} \in \mathcal{S}$ where the minimum of Equation (A.34) occurs. When Equation (A.34) has multiple solutions (multiple points on $\boldsymbol{x} \in \mathcal{S}$ forming specular paths with respect to $\boldsymbol{v}$), we select among these solutions the point that minimizes the difference $\|\tau_{\mathcal{F}}\left(\boldsymbol{v}\right) - \tau\left(\boldsymbol{x};\boldsymbol{v}\right)\|$; that is, the point that most closely matches the available Fermat pathlength measurement.

We then approximate the problem of Equation (A.33) with the simpler problem

$$\min_{\boldsymbol{V}^*} \left\|\tau_{\mathcal{F}}\left(\boldsymbol{v}\right) - \tau\left(\boldsymbol{x}^*,\boldsymbol{v}\right)\right\|^2, \tag{A.35}$$

where

$$\boldsymbol{x}^* = \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{T}^*} \left\| 1 - \left\langle \frac{\boldsymbol{v}-\boldsymbol{x}}{\|\boldsymbol{v}-\boldsymbol{x}\|}, \hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \right\rangle \right\|^2, \tag{A.36}$$

and $\boldsymbol{V}^*$ are the vertices of triangle $\mathcal{T}^*$. The difficulty in solving the optimization problem of Equation (A.35) comes from the fact that the point $\boldsymbol{x}^*$ is defined *implicitly* as a function of the vertices $\boldsymbol{V}^*$, through the second optimization problem of Equation (A.36). Nonetheless, we can use the *implicit function theorem* [166] to compute the derivative of $\boldsymbol{x}^*$ with respect to the vertex coordinates $\boldsymbol{V}^*$. This derivative is given by Jakob and Marschner [81], and in our case, we use automatic differentiation to compute the corresponding Jacobian terms. Given this, we can optimize Equation (A.35) using gradient descent.

In using the above procedure, we have made a number of assumptions:

- We assume that the available Fermat pathlength measurement $\tau_{\mathcal{F}}\left(\boldsymbol{v}\right)$ is a *specular* pathlength, when it can also be a boundary pathlength. In practice, we use the following heuristic to remove measurements that are likely to be boundary pathlengths: We discard any measurements for which the forward projection prob-

lem of Equation (A.34) does not have a good solution (loss function below some threshold).

- We identify the measurement $\tau_{\mathcal{F}}(\boldsymbol{v})$ with the point $\boldsymbol{x} \in \mathcal{S}$, among the solutions of the forward projection problem, whose pathlength $\tau(\boldsymbol{x}; \boldsymbol{v})$ is the closest to the measurement. It is possible that $\tau_{\mathcal{F}}(\boldsymbol{v})$ may correspond to a different specular point. In practice, we found this heuristic to work well.
- When approximating Equation (A.33) with Equation (A.35), we do not take into account that, as the surface deforms, the triangle $\mathcal{T}^*$ that contains the point $\boldsymbol{x}$ producing the specular path may change. In practice, we address this by alternating between the optimization problems of Equations (A.35) and (A.34).

Finally, we mention that when we have more than one Fermat pathlength measurements $\{\tau_{\mathcal{F}}(\boldsymbol{v}_m), m = 1, \ldots, M\}$, the same vertex in $\boldsymbol{V}$ can be used by more than one triangles $\mathcal{T}_m^*$, each corresponding to the solution of the forward projection for measurement $\tau_{\mathcal{F}}(\boldsymbol{v}_m)$. In that case, during optimization, each vertex is updated by the sum of the gradients for all problems (A.35) affecting it.

A.6 Proof of Proposition 6′

Proposition 6′. Fermat's principle. *Let $(p, q) \in [0, 1]^2$ be a parameterization of the surface $\mathcal{X}$. Then, for any pair of points $\boldsymbol{v}_s$ and $\boldsymbol{v}_d$,*

$$\mathcal{S}(\boldsymbol{v}_s, \boldsymbol{v}_d) = \left\{ \boldsymbol{x} \in \mathcal{X} : \nabla_{(p,q)} \tau(\boldsymbol{x}(p, q); \boldsymbol{v}_s, \boldsymbol{v}_d) = \boldsymbol{0} \right\}. \tag{A.37}$$

Let $r \in [0, 1]$ be a parameterization of the surface boundary $\partial\mathcal{X}$. Then, for any pair of points $\boldsymbol{v}_s$ and $\boldsymbol{v}_d$,

$$\mathcal{B}(\boldsymbol{v}_s, \boldsymbol{v}_d) = \{ \boldsymbol{x} \in \partial\mathcal{X} : \partial\tau(\boldsymbol{x}(r); \boldsymbol{v}_s, \boldsymbol{v}_d) / \partial r = 0 \}. \tag{A.38}$$

Proof. For the first part of the proposition, Equation (A.37), we have,

$$\frac{\partial \tau(\boldsymbol{x}(p, q); \boldsymbol{v}_s, \boldsymbol{v}_d)}{\partial p} = 2 \cdot \left\langle \frac{\boldsymbol{x}(p, q) - \boldsymbol{v}_s}{\|\boldsymbol{x}(p, q) - \boldsymbol{v}_s\|} + \frac{\boldsymbol{x}(p, q) - \boldsymbol{v}_d}{\|\boldsymbol{x}(p, q) - \boldsymbol{v}_d\|}, \boldsymbol{x}_p(p, q) \right\rangle = s \left\langle \boldsymbol{h}(\boldsymbol{x}(p, q); \boldsymbol{v}_s, \boldsymbol{v}_d), \boldsymbol{x}_p(p, q) \right\rangle \tag{A.39}$$

where $\boldsymbol{x}_p(p, q)$ is the partial derivative of $\boldsymbol{x}$ with respect to p, $\boldsymbol{h}(\boldsymbol{x}(p, q); \boldsymbol{v}_s, \boldsymbol{v}_d)$ is the half-vector corresponding to the directions parallel to $\boldsymbol{x}(p, q) - \boldsymbol{v}_s$ and $\boldsymbol{x}(p, q) - \boldsymbol{v}_d$, and

s some scalar. The vector $\boldsymbol{x}_p(p,q)$ is tangent to the surface $\mathcal{X}$ at $\boldsymbol{x}$, and therefore orthogonal to the surface normal $\hat{\boldsymbol{n}}(\boldsymbol{x})$ at that point. If $\boldsymbol{x}(p,q) \in \mathcal{S}(\boldsymbol{v}_s, \boldsymbol{v}_d)$, then from the specular reflection property, the half-vector $\boldsymbol{h}(\boldsymbol{x}(p,q); \boldsymbol{v}_s, \boldsymbol{v}_d)$ is parallel to the normal $\hat{\boldsymbol{n}}(\boldsymbol{x})$. Therefore,

$$\boldsymbol{h}(\boldsymbol{x}(p,q); \boldsymbol{v}_s, \boldsymbol{v}_d) \parallel \hat{\boldsymbol{n}}(\boldsymbol{x}) \text{ and } \boldsymbol{x}_p(p,q) \perp \hat{\boldsymbol{n}}(\boldsymbol{x}) \Rightarrow \left\langle \boldsymbol{h}(\boldsymbol{x}(p,q); \boldsymbol{v}_s, \boldsymbol{v}_d), \boldsymbol{x}_p(p,q) \right\rangle = 0, \tag{A.40}$$

and from Equation (A.39),

$$\frac{\partial \tau(\boldsymbol{x}(p,q); \boldsymbol{v}_s, \boldsymbol{v}_d)}{\partial p} = 0. \tag{A.41}$$

The proof for $\frac{\partial \tau(\boldsymbol{x}(p,q); \boldsymbol{v}_s, \boldsymbol{v}_d)}{\partial q} = 0$ is exactly the same. Therefore, if $\boldsymbol{x}(p,q) \in \mathcal{S}(\boldsymbol{v}_s, \boldsymbol{v}_d)$, then $\nabla_{(p,q)} \tau(\boldsymbol{x}(p,q); \boldsymbol{v}_s, \boldsymbol{v}_d) = \boldsymbol{0}$. Conversely, if $\nabla_{(p,q)} \tau(\boldsymbol{x}(p,q); \boldsymbol{v}_s, \boldsymbol{v}_d) = \boldsymbol{0}$, from Equation (A.39) we have that either $\boldsymbol{h}(\boldsymbol{x}(p,q); \boldsymbol{v}_s, \boldsymbol{v}_d) = \boldsymbol{0}$, or that $\boldsymbol{h}(\boldsymbol{x}(p,q); \boldsymbol{v}_s, \boldsymbol{v}_d)$ is orthogonal to both $\boldsymbol{x}_p(p,q)$ and $\boldsymbol{x}_q(p,q)$, and therefore parallel to the normal $\hat{\boldsymbol{n}}(\boldsymbol{x})$. By assuming that the points $\boldsymbol{v}_s$ and $\boldsymbol{v}_d$ must be on the same side of the scene $\mathcal{X}$, this implies that $\boldsymbol{x}(p,q) \in \mathcal{S}(\boldsymbol{v}_s, \boldsymbol{v}_d)$. This concludes the proof of the first part of the proposition.

For the second part of the proposition, Equation (A.38), we have,

$$\frac{\partial \tau(\boldsymbol{x}(r); \boldsymbol{v}_s, \boldsymbol{v}_d)}{\partial r} = 2 \cdot \left\langle \frac{\boldsymbol{x}(r) - \boldsymbol{v}_s}{\|\boldsymbol{x}(r) - \boldsymbol{v}_s\|} + \frac{\boldsymbol{x}(r) - \boldsymbol{v}_d}{\|\boldsymbol{x}(r) - \boldsymbol{v}_d\|}, \boldsymbol{x}_p(r) \right\rangle = s \left\langle \boldsymbol{h}(\boldsymbol{x}(r); \boldsymbol{v}_s, \boldsymbol{v}_d), \boldsymbol{x}_p(r) \right\rangle \tag{A.42}$$

where $\boldsymbol{x}_r(r)$ is the partial derivative of $\boldsymbol{x}$ with respect to r. The vector $\boldsymbol{x}_r(r)$ is parallel to the tangent $\hat{\boldsymbol{t}}(\boldsymbol{x})$ of the curve $\partial\mathcal{X}$ at $\boldsymbol{x}$. If $\boldsymbol{x}(r) \in \mathcal{B}(\boldsymbol{v})$, then from the property of boundary Fermat paths, the half-vector $\boldsymbol{h}(\boldsymbol{x}(r); \boldsymbol{v}_s, \boldsymbol{v}_d)$ is orthogonal to the tangent $\hat{\boldsymbol{t}}(\boldsymbol{x})$. Therefore,

$$\boldsymbol{h}(\boldsymbol{x}(r), \boldsymbol{v}_s, \boldsymbol{v}_d) \perp \hat{\boldsymbol{t}}(\boldsymbol{x}) \text{ and } \boldsymbol{x}_r(r) \parallel \hat{\boldsymbol{t}}(\boldsymbol{x}) \Rightarrow \left\langle \boldsymbol{h}(\boldsymbol{x}(p,q); \boldsymbol{v}_s, \boldsymbol{v}_d), \boldsymbol{x}_r(r) \right\rangle = 0, \tag{A.43}$$

and from Equation (A.42),

$$\frac{\partial \tau(\boldsymbol{x}(r); \boldsymbol{v}_s, \boldsymbol{v}_d)}{\partial r} = 0. \tag{A.44}$$

Conversely, if $\frac{\partial \tau(\boldsymbol{x}(r); \boldsymbol{v}_s, \boldsymbol{v}_d)}{\partial r} = 0$, from Equation (A.39) we have that either $\boldsymbol{h}(\boldsymbol{x}(p,q); \boldsymbol{v}_s, \boldsymbol{v}_d) = \boldsymbol{0}$, or that $\boldsymbol{h}(\boldsymbol{x}(p,q); \boldsymbol{v}_s, \boldsymbol{v}_d)$ is orthogonal to $\boldsymbol{x}_r(r)$, and therefore orthogonal to $\hat{\boldsymbol{t}}(\boldsymbol{x})$. By assuming that the points $\boldsymbol{v}_s$ and $\boldsymbol{v}_d$ are at the same side of the boundary $\partial\mathcal{X}$, this implies that $\boldsymbol{x}(r) \in \mathcal{B}(\boldsymbol{v}_s, \boldsymbol{v}_d)$. This concludes the proof of the second part of the proposition. □

A.7 Proof of Proposition 8′

Proposition 8′. Fermat flow equation. *Consider a branch of the Fermat pathlength function* $\tau_{\mathcal{F}}(\boldsymbol{v}_s, \boldsymbol{v}_d)$ *corresponding to points* $\boldsymbol{v}_s, \boldsymbol{v}_d$*. Assume that there is a unique point* $\boldsymbol{x}_{\mathcal{F}} \in \mathcal{F}(\boldsymbol{v}_s, \boldsymbol{v}_d)$ *with* $\tau(\boldsymbol{x}_{\mathcal{F}}; \boldsymbol{v}_s, \boldsymbol{v}_d) = \tau_{\mathcal{F}}(\boldsymbol{v}_s, \boldsymbol{v}_d)$*. Then,*

$$\nabla_{\boldsymbol{v}_s} \tau_{\mathcal{F}}(\boldsymbol{v}_s, \boldsymbol{v}_d) = -\frac{\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}_s}{\|\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}_s\|} = -\hat{\boldsymbol{w}}_{\mathcal{F}}(\boldsymbol{v}_s, \boldsymbol{v}_d)\ , \qquad \textit{and} \tag{A.45}$$

$$\nabla_{\boldsymbol{v}_d} \tau_{\mathcal{F}}(\boldsymbol{v}_s, \boldsymbol{v}_d) = -\frac{\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}_d}{\|\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}_d\|} = -\hat{\boldsymbol{w}}_{\mathcal{F}}(\boldsymbol{v}_s, \boldsymbol{v}_d)\ . \tag{A.46}$$

Proof. We will prove Equation (A.46) only for $\nabla_{\boldsymbol{v}_s} \tau_{\mathcal{F}}(\boldsymbol{v}_s, \boldsymbol{v}_d)$, as the proof for $\nabla_{\boldsymbol{v}_d} \tau_{\mathcal{F}}(\boldsymbol{v}_s, \boldsymbol{v}_d)$ is exactly analogous. We will be using $\boldsymbol{v}_s = \left[v_s^x, v_s^y, v_s^z\right]^T$ to denote the 3D coordinates of the point $\boldsymbol{v}_s$, and similarly for all other vectors.

We first prove the proposition for the case of a *specular* path, that is, $\boldsymbol{x}_{\mathcal{F}} \in \mathcal{S}(\boldsymbol{v}_s, \boldsymbol{v}_d)$. Let $(p, q) \in [0,1]^2$ be a parameterization of the surface $\mathcal{X}$ in a neighborhood $\mathcal{N} \subset \mathcal{X}$ around $\boldsymbol{x}_{\mathcal{F}}$, such that $\boldsymbol{x}_{\mathcal{F}} = \boldsymbol{x}(p(\boldsymbol{v}_s, \boldsymbol{v}_d), q(\boldsymbol{v}_s, \boldsymbol{v}_d))$.

We use $\mathcal{O}$ to denote the *orthotomic* surface corresponding to $\mathcal{N}$ with respect to the point $\boldsymbol{v}_d$ [26, 27]. For each pair of points $\boldsymbol{v}_s$ and $\boldsymbol{x}_{\mathcal{F}} \in \mathcal{S}(\boldsymbol{v}_s, \boldsymbol{v}_d)$, the orthotomic contains a corresponding point $\boldsymbol{o}_{\mathcal{F}} = \boldsymbol{o}(p(\boldsymbol{v}_s, \boldsymbol{v}_d), q(\boldsymbol{v}_s, \boldsymbol{v}_d))$ such that

$$\|\boldsymbol{o}_{\mathcal{F}} - \boldsymbol{v}_s\| = \tau_{\mathcal{F}}, \qquad \text{and} \qquad \hat{\boldsymbol{n}}(\boldsymbol{o}_{\mathcal{F}}) = -\frac{\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}_s}{\|\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}_s\|} = -\frac{\boldsymbol{o}_{\mathcal{F}} - \boldsymbol{v}_s}{\|\boldsymbol{o}_{\mathcal{F}} - \boldsymbol{v}_s\|}, \tag{A.47}$$

where $\hat{\boldsymbol{n}}(\boldsymbol{o}_{\mathcal{F}})$ is the normal of the orthotomic $\mathcal{O}$ at $\boldsymbol{o}_{\mathcal{F}}$. Additionally, the parameterization (p, q) can be used to parameterize the orthotomic as well, through the mapping from points of the neighborhood $\mathcal{N}$ to $\mathcal{O}$. For more about the properties of the orthotomic, we refer to [26, 27].

We now consider each coordinate of the vector $\nabla_{\boldsymbol{v}_s} \tau_{\mathcal{F}}(\boldsymbol{v}_s, \boldsymbol{v}_d)$ separately. For the first coordinate, and using Equation (A.47), we have

$$\frac{\partial \tau_{\mathcal{F}}(\boldsymbol{v}_s, \boldsymbol{v}_d)}{\partial v_s^x} = \frac{\partial \|\boldsymbol{o}_{\mathcal{F}} - \boldsymbol{v}_s\|}{\partial v_s^x} \tag{A.48}$$

$$= \frac{\partial \|\boldsymbol{o}(p(\boldsymbol{v}_s, \boldsymbol{v}_d), q(\boldsymbol{v}_s, \boldsymbol{v}_d)) - \boldsymbol{v}_s\|}{\partial v^x} \tag{A.49}$$

$$= \left\langle \frac{\boldsymbol{o}(p(\boldsymbol{v}_s, \boldsymbol{v}_d), q(\boldsymbol{v}_s, \boldsymbol{v}_d)) - \boldsymbol{v}_s}{\|\boldsymbol{o}(p(\boldsymbol{v}_s, \boldsymbol{v}_d), q(\boldsymbol{v}_s, \boldsymbol{v}_d)) - \boldsymbol{v}_s\|}, \frac{\partial (\boldsymbol{o}(p(\boldsymbol{v}_s, \boldsymbol{v}_d), q(\boldsymbol{v}_s, \boldsymbol{v}_d)) - \boldsymbol{v}_s)}{\partial v_s^x} \right\rangle \tag{A.50}$$

$$= \left\langle \frac{\boldsymbol{o}_{\mathcal{F}} - \boldsymbol{v}_s}{\|\boldsymbol{o}_{\mathcal{F}} - \boldsymbol{v}_s\|}, \boldsymbol{o}_p(p(\boldsymbol{v}_s, \boldsymbol{v}_d), q(\boldsymbol{v}_s, \boldsymbol{v}_d)) \frac{\partial p(\boldsymbol{v}_s, \boldsymbol{v}_d)}{\partial v_s^x} + \boldsymbol{o}_q(p(\boldsymbol{v}_s, \boldsymbol{v}_d), q(\boldsymbol{v}_s, \boldsymbol{v}_d)) \frac{\partial q(\boldsymbol{v}_s, \boldsymbol{v}_d)}{\partial v_s^x} \right. \tag{A.51}$$

In Equation (A.51), we can use Equation (A.47) to replace $\hat{\boldsymbol{n}}\left(\boldsymbol{o}_{\mathcal{F}}\right)$, which gives us

$$\begin{aligned}\frac{\partial \tau_{\mathcal{F}}\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)}{\partial v_s^x} = & -\left\langle \hat{\boldsymbol{n}}\left(\boldsymbol{o}_{\mathcal{F}}\right), \boldsymbol{o}_p\left(p\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right), q\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)\right)\right\rangle \frac{\partial p\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)}{\partial v_s^x} \\ & -\left\langle \hat{\boldsymbol{n}}\left(\boldsymbol{o}_{\mathcal{F}}\right), \boldsymbol{o}_q\left(p\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right), q\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)\right)\right\rangle \frac{\partial q\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)}{\partial v_s^x} \\ & -\left\langle \hat{\boldsymbol{n}}\left(\boldsymbol{o}_{\mathcal{F}}\right), [1,0,0]^T\right\rangle . \end{aligned} \tag{A.52}$$

In Equation (A.52), the vectors $\boldsymbol{o}_p\left(p\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right), q\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)\right)$ and $\boldsymbol{o}_q\left(p\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right), q\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)\right)$ are tangent to the orthotomic surface $\mathcal{O}$ at $\boldsymbol{o}_{\mathcal{F}}$, and therefore orthogonal to the normal $\hat{\boldsymbol{n}}\left(\boldsymbol{o}_{\mathcal{F}}\right)$ at that point. Consequently, and using Equation (A.47), Equation (A.52) becomes

$$\frac{\partial \tau_{\mathcal{F}}\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)}{\partial v_s^x} = -0 \cdot \frac{\partial p\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)}{\partial v_s^x} - 0 \cdot \frac{\partial q\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)}{\partial v_s^x} - \left(\hat{\boldsymbol{n}}\left(\boldsymbol{o}_{\mathcal{F}}\right)\right)^x = -\frac{\left(\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}_s\right)^x}{\left\|\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}_s\right\|}. \tag{A.53}$$

Exactly analogously, we can prove that

$$\frac{\partial \tau_{\mathcal{F}}\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)}{\partial v_s^y} = -\frac{\left(\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}_s\right)^y}{\left\|\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}_s\right\|} \qquad \text{and} \qquad \frac{\partial \tau_{\mathcal{F}}\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)}{\partial v_s^z} = -\frac{\left(\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}_s\right)^z}{\left\|\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}_s\right\|}. \tag{A.54}$$

Combining Equations (A.53) and (A.54) completes the proof for the specular case.

We now prove the proposition for the case of a *boundary* path, that is, $\boldsymbol{x}_{\mathcal{F}} \in \mathcal{B}\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)$. Let $r \in [0,1]$ be a parameterization of the surface boundary $\partial\mathcal{X}$ in a neighborhood $\mathcal{M} \subset \partial\mathcal{X}$ around $\boldsymbol{x}_{\mathcal{F}}$, such that $\boldsymbol{x}_{\mathcal{F}} = \boldsymbol{x}\left(r\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)\right)$.

As in the specular case, we consider the orthotomic curve $\mathcal{Q}$ of $\mathcal{M}$ with respect to the point $\boldsymbol{v}_d$ [26, 27]. For each pair of points $\boldsymbol{v}_s$ and $\boldsymbol{x}_{\mathcal{F}} \in \mathcal{B}\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)$, the orthotomic contains a corresponding point $\boldsymbol{q}_{\mathcal{F}} = \boldsymbol{q}\left(r\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)\right)$ such that

$$\left\|\boldsymbol{q}_{\mathcal{F}} - \boldsymbol{v}_s\right\| = \tau_{\mathcal{F}}, \qquad \text{and} \qquad \hat{\boldsymbol{t}}\left(\boldsymbol{q}_{\mathcal{F}}\right) \perp -\frac{\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}_s}{\left\|\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}_s\right\|} = -\frac{\boldsymbol{q}_{\mathcal{F}} - \boldsymbol{v}_s}{\left\|\boldsymbol{q}_{\mathcal{F}} - \boldsymbol{v}_s\right\|}, \tag{A.55}$$

where $\hat{\boldsymbol{t}}\left(\boldsymbol{q}_{\mathcal{F}}\right)$ is the tangent to the orthotomic $\mathcal{Q}$ at $\boldsymbol{q}_{\mathcal{F}}$. Additionally, the parameterization r can be used to parameterize the orthotomic curve as well, through the mapping from points of the neighborhood $\mathcal{M}$ to $\mathcal{Q}$.

We now consider each coordinate of the vector $\nabla_{\boldsymbol{v}_s} \tau_{\mathcal{F}}\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)$ separately. For the first

coordinate, and using Equation (A.55), we have

$$\frac{\partial \tau_{\mathcal{F}}(\boldsymbol{v}_s, \boldsymbol{v}_d)}{\partial v_s^x} = \frac{\partial \left\| \boldsymbol{q}_{\mathcal{F}} - \boldsymbol{v}_s \right\|}{\partial v_s^x} \tag{A.56}$$

$$= \frac{\partial \left\| \boldsymbol{q}\left(r\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)\right) - \boldsymbol{v}_s \right\|}{\partial v^x} \tag{A.57}$$

$$= \left\langle \frac{\boldsymbol{q}\left(r\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)\right) - \boldsymbol{v}_s}{\left\| \boldsymbol{q}\left(r\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)\right) - \boldsymbol{v}_s \right\|}, \frac{\partial \left(\boldsymbol{q}\left(r\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)\right) - \boldsymbol{v}_s\right)}{\partial v_s^x} \right\rangle \tag{A.58}$$

$$= \left\langle \frac{\boldsymbol{q}_{\mathcal{F}} - \boldsymbol{v}_s}{\left\| \boldsymbol{q}_{\mathcal{F}} - \boldsymbol{v}_s \right\|}, \boldsymbol{q}_r\left(r\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)\right) \frac{\partial r\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)}{\partial v_s^x} - [1,0,0]^T \right\rangle \tag{A.59}$$

$$= \left\langle \frac{\boldsymbol{q}_{\mathcal{F}} - \boldsymbol{v}_s}{\left\| \boldsymbol{q}_{\mathcal{F}} - \boldsymbol{v}_s \right\|}, \boldsymbol{q}_r\left(r\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)\right) \right\rangle \frac{\partial r\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)}{\partial v_s^x} - \left\langle \frac{\boldsymbol{q}_{\mathcal{F}} - \boldsymbol{v}_s}{\left\| \boldsymbol{q}_{\mathcal{F}} - \boldsymbol{v}_s \right\|}, [1,0,0]^T \right\rangle. \tag{A.60}$$

In Equation (A.60), the vector $\boldsymbol{q}_r\left(r\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)\right)$ is parallel to the tanget $\hat{\boldsymbol{t}}\left(\boldsymbol{q}_{\mathcal{F}}\right)$ of the orthotomic $\mathcal{Q}$ at $\boldsymbol{q}_{\mathcal{F}}$. Therefore, using Equation (A.55), we can write

$$\frac{\partial \tau_{\mathcal{F}}(\boldsymbol{v}_s, \boldsymbol{v}_d)}{\partial v_s^x} = 0 \frac{\partial r\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)}{\partial v_s^x} - \left\langle \frac{\boldsymbol{q}_{\mathcal{F}} - \boldsymbol{v}_s}{\left\| \boldsymbol{q}_{\mathcal{F}} - \boldsymbol{v}_s \right\|}, [1,0,0]^T \right\rangle = -\frac{\left(\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}_s\right)^x}{\left\| \boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}_s \right\|} \tag{A.61}$$

Exactly analogously, we can prove that

$$\frac{\partial \tau_{\mathcal{F}}(\boldsymbol{v}_s, \boldsymbol{v}_d)}{\partial v_s^y} = -\frac{\left(\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}_s\right)^y}{\left\| \boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}_s \right\|} \quad \text{and} \quad \frac{\partial \tau_{\mathcal{F}}(\boldsymbol{v}_s, \boldsymbol{v}_d)}{\partial v_s^z} = -\frac{\left(\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}_s\right)^z}{\left\| \boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}_s \right\|}. \tag{A.62}$$

Combining Equations (A.61) and (A.62) completes the proof for the boundary case. □

A.8 Proof of Proposition 10′

Proposition 10′. *Let a point $\boldsymbol{x}_{\mathcal{S}} \in \mathcal{S}\left(\boldsymbol{v}_s, \boldsymbol{v}_d\right)$ belong to the specular set. Let $E\left(\tau_{\mathcal{S}}; \boldsymbol{v}_s, \boldsymbol{v}_d\right)$ be the corresponding osculating ellipsoid. If $\kappa_{min}, \kappa_{max}$ are the principal curvatures of $\mathcal{X}$ at $\boldsymbol{x}_{\mathcal{S}}$, and $\lambda_{min}, \lambda_{max}$ the principal curvatures of $E\left(\tau_{\mathcal{S}}; \boldsymbol{v}_s, \boldsymbol{v}_d\right)$ at $\boldsymbol{x}_{\mathcal{S}}$, then:*

- *If $\tau_{\mathcal{S}}$ is a local minimum of $\tau\left(\boldsymbol{x}; \boldsymbol{v}_d, \boldsymbol{v}_s\right)$, then $\lambda_{min} \leq \kappa_{min}$.*
- *If $\tau_{\mathcal{S}}$ is a local maximum of $\tau\left(\boldsymbol{x}; \boldsymbol{v}_d, \boldsymbol{v}_s\right)$, then $\kappa_{max} \leq \lambda_{max}$.*

Proof. We consider first the case where $\tau_{\mathcal{S}}$ is a local minimum of $\tau\left(\boldsymbol{x}; \boldsymbol{v}_s, \boldsymbol{v}_d\right)$. Then, there will be some neighborhood $\mathcal{N}\left(\boldsymbol{x}_{\mathcal{S}}\right) \subset \mathcal{X}$ of $\boldsymbol{x}_{\mathcal{S}}$ on $\mathcal{X}$ such that, for all points $\boldsymbol{x} \in \mathcal{N}\left(\boldsymbol{x}_{\mathcal{S}}\right)$, $\tau\left(\boldsymbol{x}; \boldsymbol{v}_s, \boldsymbol{v}_d\right) \geq \tau_{\mathcal{S}}$. Equivalently, all of $\mathcal{N}\left(\boldsymbol{x}_{\mathcal{S}}\right)$ lies outside the ellipsoid $E\left(\tau_{\mathcal{S}}; \boldsymbol{v}_s, \boldsymbol{v}_d\right)$, and is tangent to that ellipsoid at the point $\boldsymbol{x}_{\mathcal{S}}$. Therefore, the curves on $\mathcal{N}\left(\boldsymbol{x}_{\mathcal{S}}\right)$ passing through $\boldsymbol{x}_{\mathcal{S}}$ in all possible tangent directions are also tangent to and

outside of the ellipsoid $E(\tau_{\mathcal{S}}; \boldsymbol{v}_s, \boldsymbol{v}_d)$. Consequently, all normal curvatures of $\mathcal{X}$ at $\boldsymbol{x}_{\mathcal{S}}$ are greater than the smallest normal curvature of the ellipsoid $E(\tau_{\mathcal{S}}; \boldsymbol{v}_s, \boldsymbol{v}_d)$ at $\boldsymbol{x}_{\mathcal{S}}$. From the definition of principal curvatures, we conclude that $\lambda_{\min} \leq \kappa_{\min}$.

The case when $\tau_{\mathcal{S}}$ is a local maximum of $\tau(\boldsymbol{x}; \boldsymbol{v}_s, \boldsymbol{v}_d)$ proceeds very similarly: In this case, all of $\mathcal{N}(\boldsymbol{x}_{\mathcal{S}})$ is *inside* the ellipsoid $E(\tau_{\mathcal{S}}; \boldsymbol{v}_s, \boldsymbol{v}_d)$, and tangent to it at $\boldsymbol{x}_{\mathcal{S}}$. Therefore, the curves on $\mathcal{N}(\boldsymbol{x}_{\mathcal{S}})$ passing through $\boldsymbol{x}_{\mathcal{S}}$ in all possible tangent directions are also tangent to and *inside* of the ellipsoid $E(\tau_{\mathcal{S}}; \boldsymbol{v}_s, \boldsymbol{v}_d)$. Consequently, all normal curvatures of $\mathcal{X}$ at $\boldsymbol{x}_{\mathcal{S}}$ are *smaller* than the largest normal curvature of the ellipsoid $E(\tau_{\mathcal{S}}; \boldsymbol{v}_s, \boldsymbol{v}_d)$ at $\boldsymbol{x}_{\mathcal{S}}$. From the definition of principal curvatures, we conclude that $\kappa_{\max} \leq \lambda_{\max}$.

□

A.9 Derivation of Incident and Outgoing Pathlengths (Equations (2.47) and (2.48))

Let the light path $\boldsymbol{v}_s \to \boldsymbol{x}_{\mathcal{F}} \in \mathcal{X} \to \boldsymbol{v}_d$ be a Fermat path. We use $\tau_{\mathcal{F}}^{(i)}$ and $\tau_{\mathcal{F}}^{(r)}$ to denote the incident and outgoing pathlengths, respectively,

$$\tau_{\mathcal{F}}^{(i)} = \|\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}_s\|, \text{ and } \tau_{\mathcal{F}}^{(r)} = \|\boldsymbol{v}_d - \boldsymbol{x}_{\mathcal{F}}\|. \tag{A.63}$$

The point $\boldsymbol{x}_{\mathcal{F}}$ is $\tau_{\mathcal{F}}^{(r)}$ away from $\boldsymbol{v}_d$ in the direction of $\hat{\boldsymbol{w}}_{\mathcal{F}}$:

$$\boldsymbol{x}_{\mathcal{F}} = \boldsymbol{v}_d + \tau_{\mathcal{F}}^{(r)} \hat{\boldsymbol{w}}_{\mathcal{F}}. \tag{A.64}$$

Plugging it into $\tau_{\mathcal{F}}^{(i)}$ yields

$$\tau_{\mathcal{F}}^{(i)} = \|\boldsymbol{x}_{\mathcal{F}} - \boldsymbol{v}_s\| = \left\|\boldsymbol{v}_d + \tau_{\mathcal{F}}^{(r)} \hat{\boldsymbol{w}}_{\mathcal{F}} - \boldsymbol{v}_s\right\|. \tag{A.65}$$

Noting that the total pathlength equals $\tau_{\mathcal{F}} = \tau_{\mathcal{F}}^{(i)} + \tau_{\mathcal{F}}^{(r)}$, we have

$$\left\|\boldsymbol{v}_d + \tau_{\mathcal{F}}^{(r)} \hat{\boldsymbol{w}}_{\mathcal{F}} - \boldsymbol{v}_s\right\| = \tau_{\mathcal{F}} - \tau_{\mathcal{F}}^{(r)}. \tag{A.66}$$

Taking the square of both sides and noting that $\|\hat{\boldsymbol{w}}_{\mathcal{F}}\|^2 = 1$, we get

$$\|\boldsymbol{v}_d - \boldsymbol{v}_s\|^2 + 2\left\langle \boldsymbol{v}_d - \boldsymbol{v}_s, \hat{\boldsymbol{w}}_{\mathcal{F}} \right\rangle \tau_{\mathcal{F}}^{(r)} + {\tau_{\mathcal{F}}^{(r)}}^2 = {\tau_{\mathcal{F}}}^2 - 2\tau_{\mathcal{F}}\tau_{\mathcal{F}}^{(r)} + {\tau_{\mathcal{F}}^{(r)}}^2. \tag{A.67}$$

Canceling out the second-degree terms, and rearranging the rest of the terms, we have

$$\tau_{\mathcal{F}}^{(r)} = \frac{{\tau_{\mathcal{F}}}^2 - \|\boldsymbol{v}_d - \boldsymbol{v}_s\|^2}{2\left(\tau_{\mathcal{F}} + \left\langle \boldsymbol{v}_d - \boldsymbol{v}_s, \hat{\boldsymbol{w}}_{\mathcal{F}} \right\rangle\right)}, \tag{A.68}$$

and then

$$\tau_{\mathcal{F}}^{(i)} = \tau_{\mathcal{F}} - \tau_{\mathcal{F}}^{(r)} = \frac{{\tau_{\mathcal{F}}}^2 + 2\left\langle \boldsymbol{v}_d - \boldsymbol{v}_s, \hat{\boldsymbol{w}}_{\mathcal{F}} \right\rangle \tau_{\mathcal{F}} + \left\| \boldsymbol{v}_d - \boldsymbol{v}_s \right\|^2}{2\left(\tau_{\mathcal{F}} + \left\langle \boldsymbol{v}_d - \boldsymbol{v}_s, \hat{\boldsymbol{w}}_{\mathcal{F}} \right\rangle\right)}, \tag{A.69}$$

which completes the derivation of Equations (2.47) and (2.48).

Appendix B

Appendix to Chapter 3

In this appendix, we provide more details for the proof of Proposition 11.

B.1 Proof of Proposition 11

Proposition 11. *Assume that the BRDF of the surface $\mathcal{X}$ is non-zero in the specular direction. Then, for all $\boldsymbol{x} \in \mathcal{F}(\boldsymbol{v})$, the transient $I(\tau; \boldsymbol{v})$ will have a discontinuity at pathlength $\tau(\boldsymbol{x}; \boldsymbol{v})$. If $\boldsymbol{x} \in \mathcal{S}(\boldsymbol{v})$, then $I(\tau; \boldsymbol{v})$ will additionally have a vertical asymptote at $\tau(\boldsymbol{x}; \boldsymbol{v})$.*

Proof. Let $\text{Sph}(\rho; \boldsymbol{v})$ be the sphere of center $\boldsymbol{v}$ and radius ρ. Let $\mathcal{C}(\rho; \boldsymbol{v})$ be the intersection of $\text{Sph}(\rho; \boldsymbol{v})$ with $\mathcal{X}$, parameterized by $c \in [0, 1]$. Then, we can use $(c, \rho) \in [0, 1] \times [0, \infty)$ to reparameterize $\mathcal{X}$. We note, however, that this parameterization is continuously-differentiable only *locally*, separating the surface $\mathcal{X}$ into submanifolds $\mathcal{M}_i$ within which this condition holds. These submanifolds are separated by occluding contours or surface boundaries (including surface discontinuities). We can then express the transient $I(\tau; \boldsymbol{v})$ as

$$I(\tau; \boldsymbol{v}) = \sum_{\mathcal{M}_i} \int_{\mathcal{M}_i} f(\boldsymbol{x}; \boldsymbol{v}) \, \delta(\tau - \tau(\boldsymbol{x}; \boldsymbol{v})) \left| \mathcal{J}^{(c,\rho)}_{(p,q)}(\boldsymbol{x}) \right|^{-1} \mathrm{d}A(c, \rho), \tag{B.1}$$

where (p, q) is the parameterization of $\mathcal{X}$ $\mathcal{J}^{(c,\rho)}_{(p,q)}(\boldsymbol{x})$ is the Jacobian of the transformation $(p, q) \mapsto (c, \rho)$. We also consider the transient produced by each submanifold,

$$I_{\mathcal{M}_i}(\tau; \boldsymbol{v}) = \int_{\mathcal{M}_i} f(\boldsymbol{x}; \boldsymbol{v}) \, \delta(\tau - \tau(\boldsymbol{x}; \boldsymbol{v})) \left| \mathcal{J}^{(c,\rho)}_{(p,q)}(\boldsymbol{x}) \right|^{-1} dA(c, \rho), \tag{B.2}$$

Within each submanifold, the parameter ρ has a range $\rho \in [\rho_{\min}, \rho_{\max}]$. Recognizing that at each point $\boldsymbol{x} \in \mathcal{M}_i$, $\rho(\boldsymbol{x}) = \tau(\boldsymbol{x}; \boldsymbol{v})/2$, and from the definition of the boundary and specular sets, the extrema of ρ will occur either at points on an occluding contour (in which case the corresponding value of $I_{\mathcal{M}_i}$ will be zero), at boundary points $\boldsymbol{x} \in \mathcal{B}(\boldsymbol{v})$, or at specular points $\boldsymbol{x} \in \mathcal{S}(\boldsymbol{v})$. In the boundary case, the corresponding transient $I_{\mathcal{M}_i}$ will be discontinuous at $\tau = 2\rho$:

- If ρ is a maximum, then $I_{\mathcal{M}_i} > 0$ as $\tau \to 2\rho^-$, and $I_{\mathcal{M}_i} = 0$ as $\tau \to 2\rho^+$.
- If ρ is a minimum, then $I_{\mathcal{M}_i} > 0$ as $\tau \to 2\rho^+$, and $I_{\mathcal{M}_i} = 0$ as $\tau \to 2\rho^-$.

Consequently, the transient I will also have a discontinuity at the same point.

We now consider the particular case of a point $\boldsymbol{x}_{\mathcal{S}} \in \mathcal{S}(\boldsymbol{v})$. Recognizing that $\rho(\boldsymbol{x}_{\mathcal{S}}) = \tau(\boldsymbol{x}_{\mathcal{S}}; \boldsymbol{v})/2$, we have from Equation (2.20) that $\nabla_{(p,q)}\rho(\boldsymbol{x}_{\mathcal{S}}) = \boldsymbol{0}$. Consequently,

$$\left|\mathcal{J}^{(c,\rho)}_{(p,q)}(\boldsymbol{x}_{\mathcal{S}})\right| = \frac{\partial \rho(\boldsymbol{x}_{\mathcal{S}})}{\partial p}\frac{\partial c(\boldsymbol{x}_{\mathcal{S}})}{\partial q} - \frac{\partial \rho(\boldsymbol{x}_{\mathcal{S}})}{\partial q}\frac{\partial c(\boldsymbol{x}_{\mathcal{S}})}{\partial p} = 0. \tag{B.3}$$

Then, from Equations (B.2) and (B.3), at $\tau = \tau(\boldsymbol{x}_{\mathcal{S}}; \boldsymbol{v})$, the corresponding transient $I_{\mathcal{M}_i}$ and the total transient I converge to infinity, resulting in a discontinuity. As in the boundary case, we can distinguish whether I converges to infinity as $\tau \to 2\rho(\boldsymbol{x}_{\mathcal{S}})^-$ or $\tau \to 2\rho(\boldsymbol{x}_{\mathcal{S}})^+$ depending on whether $\tau(\boldsymbol{x}_{\mathcal{S}}; \boldsymbol{v})$ is a minimum or maximum, respectively. □

We note that the discussion in Section 3.2.1 about identifying the type of stationarity of a discontinuity follows directly from the above proof of Proposition 11.

We additionally state and prove the analogue of Proposition 11 for the non-confocal case.

Proposition 11′. *Assume that the BRDF of the surface $\mathcal{X}$ is non-zero in the specular direction. Then, for all $\boldsymbol{x} \in \mathcal{F}(\boldsymbol{v}_s, \boldsymbol{v}_d)$, the transient $I(\tau; \boldsymbol{v}_s, \boldsymbol{v}_d)$ will have a discontinuity at pathlength $\tau(\boldsymbol{x}; \boldsymbol{v}_s, \boldsymbol{v}_d)$. If $\boldsymbol{x} \in \mathcal{S}(\boldsymbol{v}_s, \boldsymbol{v}_d)$, then $I(\tau; \boldsymbol{v}_s, \boldsymbol{v}_d)$ will additionally have a local maximum at $\tau(\boldsymbol{x}; \boldsymbol{v}_s, \boldsymbol{v}_d)$.*

Proof. Let $E(\tau; \boldsymbol{v}_s, \boldsymbol{v}_d)$ be the ellipsoid of foci $\boldsymbol{v}_s$ and $\boldsymbol{v}_d$, and pathlength τ. Let $\mathcal{C}(\tau; \boldsymbol{v}_s, \boldsymbol{v}_d)$ be the intersection of $E(\tau; \boldsymbol{v}_s, \boldsymbol{v}_d)$ with $\mathcal{X}$, parameterized by $c \in [0, 1]$. Then, we can use $(c, \tau) \in [0, 1] \times [0, \infty)$ to reparameterize $\mathcal{X}$. The rest of the proof follows exactly analogously to the proof of Proposition 11 for the confocal case. □

Appendix C

Appendix to Chapter 4

In this appendix, we provide the proof of Proposition 13.

C.1 Proof of Proposition 13

Proposition 13. *Let $\mathcal{X}$ be planar. Let paths $\boldsymbol{v}_s \to \boldsymbol{x}_{\mathcal{S}1} \in \mathcal{X} \to \boldsymbol{v}_{d1}$ and $\boldsymbol{v}_s \to \boldsymbol{x}_{\mathcal{S}2} \in \mathcal{X} \to \boldsymbol{v}_{d2}$ be two specular paths, and $\tau_{\mathcal{S}1}$ and $\tau_{\mathcal{S}2}$ be the ground truth pathlengths, respectively. Assume the pathlengths $\tau_{\mathcal{S}1} + \tau_0$ and $\tau_{\mathcal{S}2} + \tau_0$ also satisfy radiometric constraints (Equation(4.13)). Plugging Equation (4.14) to Equation(4.13) leads to the following seventh-order polynomial in terms of τ_0 :*

$$\tau_0 \left(\tau_0^2 - l^2\right)^2 \left[\tau_0^2 + (m_1 + m_2)\,\tau_0 + l^2\right] = 0\,, \tag{C.1}$$

where l is twice the distance from the point light source $\boldsymbol{v}_s$ to $\mathcal{X}$, and m_1 and m_2 are defined as follows:

$$l = 2 \cdot dist\left(\boldsymbol{v}_s, \mathcal{X}\right)\ ,\ and \tag{C.2}$$

$$m_j = \tau_{\mathcal{S}j} + \left\langle \boldsymbol{v}_{dj} - \boldsymbol{v}_s, \hat{\boldsymbol{w}}_{\mathcal{S}}\left(\boldsymbol{v}_s, \boldsymbol{v}_{dj}\right)\right\rangle, j \in \{1,2\}\ . \tag{C.3}$$

Proof. We define the potentially feasible pathlengths with a global offset τ_0 as

$$\tau'_{\mathcal{S}j} = \tau_{\mathcal{S}j} + \tau_0, j \in \{1,2\}\ . \tag{C.4}$$

For a planer surface, both principal curvatures of the wavefront arriving at the camera are

$$\kappa_{1j}^{(c)} = \kappa_{2j}^{(c)} = \frac{1}{\tau_{\mathcal{S}j}}, j \in \{1,2\}\ . \tag{C.5}$$

We can then simplify the global scaled intensity (Equation (4.12)) by plugging in ground truth $\tilde{E}_j^{(c)}$ (Equation (4.14)) and the principal curvatures $\kappa_{1j}^{(c)}$:

$$\tilde{I}_s = \frac{\tilde{E}_j^{(c)} {\tau'^{(i)}_{\mathcal{S}j}}^2}{\left|\left(1-\kappa_{1j}^{(c)}\tau'^{(r)}_{\mathcal{S}j}\right)\left(1-\kappa_{2j}^{(c)}\tau'^{(r)}_{\mathcal{S}j}\right)\right|} = \frac{\tilde{E}_j^{(c)} {\tau'^{(i)}_{\mathcal{S}j}}^2}{\left(1-\kappa_{1j}^{(c)}\tau'^{(r)}_{\mathcal{S}j}\right)^2} = \frac{\frac{\tilde{I}_s}{{\tau_{\mathcal{S}j}}^2}\cdot {\tau'^{(i)}_{\mathcal{S}j}}^2}{\left(1-\frac{\tau'^{(r)}_{\mathcal{S}j}}{\tau_{\mathcal{S}j}}\right)^2} = \frac{\tilde{I}_s \cdot {\tau'^{(i)}_{\mathcal{S}j}}^2}{\left(\tau_{\mathcal{S}j}-\tau'^{(r)}_{\mathcal{S}j}\right)^2}. \tag{C.6}$$

Substituting the incident and outgoing pathlengths (Equations (2.47) and (2.48)) in Equation (C.6), we have

$$\tilde{I}_s = \frac{\tilde{I}_s \cdot \left({\tau'_{\mathcal{S}j}}^2 + 2\left\langle \boldsymbol{v}_{dj}-\boldsymbol{v}_s, \hat{\boldsymbol{w}}_{\mathcal{S}j}\right\rangle \tau'_{\mathcal{S}j} + \left\|\boldsymbol{v}_{dj}-\boldsymbol{v}_s\right\|^2\right)^2}{\left(2\left(\tau'_{\mathcal{S}j}+\left\langle \boldsymbol{v}_{dj}-\boldsymbol{v}_s, \hat{\boldsymbol{w}}_{\mathcal{S}j}\right\rangle\right)\tau_{\mathcal{S}j} - {\tau'_{\mathcal{S}j}}^2 + \left\|\boldsymbol{v}_{dj}-\boldsymbol{v}_s\right\|^2\right)^2}. \tag{C.7}$$

Ignoring $\tilde{I}_s$, the numerator $\mathcal{N}_i$ can be further expanded:

$$\mathcal{N}_i = \left(\left(\tau_{\mathcal{S}j}+\tau_0\right)^2 + 2\left\langle \boldsymbol{v}_{dj}-\boldsymbol{v}_s, \hat{\boldsymbol{w}}_{\mathcal{S}j}\right\rangle\left(\tau_{\mathcal{S}j}+\tau_0\right) + \left\|\boldsymbol{v}_{dj}-\boldsymbol{v}_s\right\|^2\right)^2 \tag{C.8}$$

$$= \left(\tau_0^2 + 2m_j\tau_0 + l_j^2\right)^2, \tag{C.9}$$

where

$$m_j = \tau_{\mathcal{S}j} + \left\langle \boldsymbol{v}_{dj}-\boldsymbol{v}_s, \hat{\boldsymbol{w}}_{\mathcal{S}j}\right\rangle, \text{ and} \tag{C.10}$$

$${l_j}^2 = {\tau_{\mathcal{S}j}}^2 + 2\left\langle \boldsymbol{v}_{dj}-\boldsymbol{v}_s, \hat{\boldsymbol{w}}_{\mathcal{S}j}\right\rangle \tau_{\mathcal{S}j} + \left\|\boldsymbol{v}_{dj}-\boldsymbol{v}_s\right\|^2. \tag{C.11}$$

One thing to note here is that for a planar surface, $l_j, j = \{1,2\}$ remains constant. This can be proved geometrically. We use $\boldsymbol{o}_{\mathcal{S}}$ to denote the mirror reflection of $\boldsymbol{v}_s$ about the plane. We define l as the distance between $\boldsymbol{o}_{\mathcal{S}}$ and $\boldsymbol{v}_s$ as:

$$l = \left\|\boldsymbol{v}_s - \boldsymbol{o}_{\mathcal{S}}\right\| = 2\cdot \text{dist}\left(\boldsymbol{v}_s, \mathcal{X}\right), j \in \{1,2\}. \tag{C.12}$$

We notice that $\tau_{\mathcal{S}} = \left\|\boldsymbol{v}_d - \boldsymbol{o}_{\mathcal{S}}\right\|$. We denote γ as the angle between vectors $\boldsymbol{v}_s - \boldsymbol{v}_d$ and $\boldsymbol{o}_{\mathcal{S}} - \boldsymbol{v}_d$. Then based on the law of cosines in trigonometry, we can relate the sides of triangle $\triangle \boldsymbol{v}_s\boldsymbol{v}_d\boldsymbol{o}_{\mathcal{S}}$ by

$$l^2 = {\tau_{\mathcal{S}}}^2 - 2\tau_{\mathcal{S}}\left\|\boldsymbol{v}_d-\boldsymbol{v}_s\right\|\cos\left(\gamma\right) + \left\|\boldsymbol{v}_d-\boldsymbol{v}_s\right\|^2$$

$$= {\tau_{\mathcal{S}}}^2 + 2\left\langle \boldsymbol{v}_d-\boldsymbol{v}_s, \hat{\boldsymbol{w}}_{\mathcal{S}}\right\rangle \tau_{\mathcal{S}} + \left\|\boldsymbol{v}_d-\boldsymbol{v}_s\right\|^2. \tag{C.13}$$

Therefore, comparing Equation (C.11) and (C.13), we have

$$l_j = l = 2 \cdot \text{dist}\left(\boldsymbol{v}_s, \mathcal{P}\right), j \in \{1, 2\} \ . \tag{C.14}$$

Similarly, for the denominator $\mathcal{D}_i$, we have

$$\begin{aligned}
\mathcal{D}_i &= \left(2\left(\tau_{\mathcal{S}j} + \tau_0 + \left\langle \boldsymbol{v}_{dj} - \boldsymbol{v}_s, \hat{\boldsymbol{w}}_{\mathcal{S}j} \right\rangle\right)\tau_{\mathcal{S}j} - \left(\tau_{\mathcal{S}j} + \tau_0\right)^2 + \left\| \boldsymbol{v}_{dj} - \boldsymbol{v}_s \right\|^2\right)^2 \\
&= \left(\tau_0^2 - l^2\right)^2 \ .
\end{aligned} \tag{C.15}$$

Finally, plugging $\mathcal{N}_i$ and $\mathcal{D}_i$ to Equation (4.13) and noticing that $\mathcal{D}_1 = \mathcal{D}_2$ and are both positive, we obtain following polynomial:

$$\begin{aligned}
&\left(\tau_0^2 - l^2\right)^2 \left[\left(\tau_0^2 + 2m_1\tau_0 + l^2\right)^2 - \left(\tau_0^2 + 2m_2\tau_0 + l^2\right)^2\right] = 0 \\
\Rightarrow\ &4\left(m_1 - m_2\right)\tau_0\left(\tau_0^2 - l^2\right)^2\left[\tau_0^2 + \left(m_1 + m_2\right)\tau_0 + l^2\right] = 0 \\
\Rightarrow\ &\tau_0\left(\tau_0^2 - l^2\right)^2\left[\tau_0^2 + \left(m_1 + m_2\right)\tau_0 + l^2\right] = 0
\end{aligned} \tag{C.16}$$

□

Appendix D

Appendix to Chapter 5

In this appendix, we cover the following topics:

1. In Section D.1, we provide the derivation of the bias correction term.
2. In Section D.2, we define the total variation function $\boldsymbol{V}(\cdot)$ and the edge map $\boldsymbol{E}$ used in the regularization terms.
3. In Section D.3, we show comparison results and ablation studies on more data in our collected Google Pixel 4 dataset.

D.1 Proof of Equation (5.5)

In this section, we provide a detailed derivation of the bias correction term. For convenience, we restate here our assumed image formation model. Given an MPI representation, its corresponding DP images can be expressed as:

$$\mathbf{I}_o^{\{l,r\}} = \mathbf{I}_b^{\{l,r\}} + \mathbf{N}^{\{l,r\}}, \tag{D.1}$$

where $\mathbf{I}_b^{\{l,r\}}$ are the latent noise-free left and right defocused images, and $\mathbf{N}^{\{l,r\}}$ is additive white Gaussian noise with entries independent identically distributed with distribution $\mathcal{N}\left(0, \sigma^2\right)$. Our goal is to optimize for an MPI with intensity-alpha layers $(\hat{\boldsymbol{c}}_i, \hat{\boldsymbol{\alpha}}_i)$, with defocus sizes $d_i, i \in [1, \ldots, N]$, such that the L_2 loss $\left\| \hat{\mathbf{I}}_b^{\{l,r\}} - \mathbf{I}_o^{\{l,r\}} \right\|_2^2$ is minimized. We show that, in the presence of image noise, minimizing the above loss biases the estimated defocus map towards smaller blur values. Specifically, we quantify this bias and then correct for it in our optimization.

For simplicity, we assume for now that all scene contents lie on a single fronto-parallel plane with ground truth defocus size $d^\star$, and our scene representation is an MPI with a single opaque layer (i.e., $\hat{\boldsymbol{\alpha}}_i = \mathbf{1}$) with a defocus size hypothesis d_i. Under this assumption, the defocused image rendering equation (Equation (5.3))

$$\hat{\mathbf{I}}_b^{\{l,r\}} = \sum_{i=1}^{N} \left[\left(\boldsymbol{k}_{d_i}^{\{l,r\}} * (\boldsymbol{c}_i \boldsymbol{\alpha}_i) \right) \odot \prod_{j=i+1}^{N} \left(\mathbf{1} - \boldsymbol{k}_{d_j}^{\{l,r\}} * \boldsymbol{\alpha}_j \right) \right] \tag{D.2}$$

reduces to

$$\hat{\mathbf{I}}_b^{\{l,r\}} = \boldsymbol{k}_{d_i}^{\{l,r\}} * \hat{\boldsymbol{c}}_i \,. \tag{D.3}$$

Similarly, Equation (D.1) becomes:

$$\mathbf{I}_o^{\{l,r\}} = \boldsymbol{k}_{d^\star}^{\{l,r\}} * \boldsymbol{c}_i + \mathbf{N}^{\{l,r\}}. \tag{D.4}$$

We can express the above equations in the frequency domain as follows:

$$\boldsymbol{\mathcal{I}}_o^{\{l,r\}} = \boldsymbol{K}_{d^\star}^{\{l,r\}} \boldsymbol{C}_i + \boldsymbol{\mathcal{N}}^{\{l,r\}} \,, \tag{D.5}$$

where $\boldsymbol{\mathcal{I}}_o^{\{l,r\}}, \boldsymbol{K}_{d^\star}^{\{l,r\}}, \boldsymbol{C}_i$, and $\boldsymbol{\mathcal{N}}^{\{l,r\}}$ are the Fourier transforms of $\mathbf{I}_o^{\{l,r\}}, \boldsymbol{k}_{d^\star}^{\{l,r\}}, \boldsymbol{c}_i$, and $\mathbf{N}^{\{l,r\}}$, respectively. Note that the entries of $\boldsymbol{\mathcal{N}}^{\{l,r\}}$ are also independent identically distributed with the same Gaussian distribution $\mathcal{N}\left(0, \sigma^2\right)$ as the entries of $\mathbf{N}^{\{l,r\}}$.

We can obtain a maximum a posteriori (MAP) estimate of $\hat{\boldsymbol{C}}_i$ and d_i by solving the following optimization problem [194]:

$$\begin{aligned} &\arg\max P\left(\boldsymbol{\mathcal{I}}_o^l, \boldsymbol{\mathcal{I}}_o^r | \hat{\boldsymbol{C}}_i, d_i, \sigma\right) P\left(\hat{\boldsymbol{C}}_i, d_i\right) \\ &= \arg\max P\left(\boldsymbol{\mathcal{I}}_o^l, \boldsymbol{\mathcal{I}}_o^r | \hat{\boldsymbol{C}}_i, d_i, \sigma\right) P\left(\hat{\boldsymbol{C}}_i\right) \,. \end{aligned} \tag{D.6}$$

According to Equation (D.5), we have

$$P\left(\boldsymbol{\mathcal{I}}_o^l, \boldsymbol{\mathcal{I}}_o^r | \hat{\boldsymbol{C}}_i, d_i, \sigma\right) \propto \exp\left(-\frac{1}{2\sigma^2} \sum_{v=\{l,r\}} \left\| \boldsymbol{K}_{d_i}^v \hat{\boldsymbol{C}}_i - \boldsymbol{\mathcal{I}}_o^v \right\|^2 \right) \,. \tag{D.7}$$

We also follow Zhou et al. [194] in assuming a prior for the latent all-in-focus image such that:

$$P\left(\hat{\boldsymbol{C}}_i\right) \propto \exp\left(-\frac{1}{2} \left\| \boldsymbol{\Phi} \hat{\boldsymbol{C}}_i \right\|^2 \right) \,, \tag{D.8}$$

where we define $\boldsymbol{\Phi}$ such that

$$\left| \boldsymbol{\Phi}\left(f\right) \right|^2 = \frac{1}{\left| \hat{\boldsymbol{C}}_i\left(f\right) \right|^2} \,, \tag{D.9}$$

and f is the frequency. As $\hat{\boldsymbol{C}}_i$ is the unknown variable, we approximate Equation (D.9) by averaging the power spectrum over a set of natural images $\{\boldsymbol{C}_i\}$:

$$|\boldsymbol{\Phi}(f)|^2 = \frac{1}{\int_{\boldsymbol{C}_i} |\boldsymbol{C}_i(f)|^2 \mu(\boldsymbol{C}_i)}, \tag{D.10}$$

where $\mu(\boldsymbol{C}_i)$ represents the probability distribution of $\boldsymbol{C}_i$ in image domain.

Maximizing the log-likelihood of Equation (D.6) is equivalent to minimizing the following loss:

$$E\left(d_i|\boldsymbol{\mathcal{I}}_o^l, \boldsymbol{\mathcal{I}}_o^r, \sigma\right) = \min_{\hat{\boldsymbol{C}}_i} \left(\sum_{v=\{l,r\}} \left\| \boldsymbol{K}_{d_i}^v \hat{\boldsymbol{C}}_i - \boldsymbol{\mathcal{I}}_o^v \right\|^2 \right) + \left\| \sigma \boldsymbol{\Phi} \hat{\boldsymbol{C}}_i \right\|^2. \tag{D.11}$$

d_i can be estimated as the minimizer of the above energy function. Then given d_i, setting $\partial E / \partial \hat{\boldsymbol{C}}_i = 0$ yields the following solution of $\hat{\boldsymbol{C}}_i$, known as a *generalized Wiener deconvolution with two observations*:

$$\hat{\boldsymbol{C}}_i = \frac{\boldsymbol{\mathcal{I}}_o^l \overline{\boldsymbol{K}_{d_i}^l} + \boldsymbol{\mathcal{I}}_o^r \overline{\boldsymbol{K}_{d_i}^r}}{\left|\boldsymbol{K}_{d_i}^l\right|^2 + \left|\boldsymbol{K}_{d_i}^r\right|^2 + \sigma^2 |\boldsymbol{\Phi}|^2}, \tag{D.12}$$

where $\overline{\boldsymbol{K}_{d_i}^{\{l,r\}}}$ is the complex conjugate of $\boldsymbol{K}_{d_i}^{\{l,r\}}$, and $\left|\boldsymbol{K}_{d_i}^{\{l,r\}}\right|^2 = \boldsymbol{K}_{d_i}^{\{l,r\}} \overline{\boldsymbol{K}_{d_i}^{\{l,r\}}}$.

We then evaluate the defocus size hypothesis d_i by computing the minimization loss given the latent ground truth depth $d^\star$, and the noise level σ, that is,

$$E\left(d_i|\boldsymbol{K}_{d^\star}^l, \boldsymbol{K}_{d^\star}^r, \sigma\right) = \mathbb{E}_{\boldsymbol{C}_i, \boldsymbol{\mathcal{I}}_o^l, \boldsymbol{\mathcal{I}}_o^r} E\left(d_i|\boldsymbol{K}_{d^\star}^l, \boldsymbol{K}_{d^\star}^r, \sigma, \boldsymbol{C}_i, \boldsymbol{\mathcal{I}}_o^l, \boldsymbol{\mathcal{I}}_o^r\right) \tag{D.13}$$

$$= \mathbb{E}_{\boldsymbol{C}_i, \boldsymbol{\mathcal{I}}_o^l, \boldsymbol{\mathcal{I}}_o^r} \left[\left(\sum_{v=\{l,r\}} \left\| \boldsymbol{K}_{d_i}^v \hat{\boldsymbol{C}}_i - \boldsymbol{\mathcal{I}}_o^v \right\|^2 \right) + \left\| \sigma \boldsymbol{\Phi} \hat{\boldsymbol{C}}_i \right\|^2 \right], \tag{D.14}$$

where $\mathbb{E}(\cdot)$ is the expectation. Substituting $\hat{\boldsymbol{C}}_i$ with Equation (D.12) gives us:

$$\begin{aligned} &E\left(d_i|\boldsymbol{K}_{d^\star}^l, \boldsymbol{K}_{d^\star}^r, \sigma\right) \\ =&\mathbb{E}_{\boldsymbol{C}_i, \boldsymbol{\mathcal{I}}_o^l, \boldsymbol{\mathcal{I}}_o^r} \left[\left(\sum_{v=\{l,r\}} \left\| \boldsymbol{K}_{d_i}^v \frac{\boldsymbol{\mathcal{I}}_o^l \overline{\boldsymbol{K}_{d_i}^l} + \boldsymbol{\mathcal{I}}_o^r \overline{\boldsymbol{K}_{d_i}^r}}{\left|\boldsymbol{K}_{d_i}^l\right|^2 + \left|\boldsymbol{K}_{d_i}^r\right|^2 + \sigma^2 |\boldsymbol{\Phi}|^2} - \boldsymbol{\mathcal{I}}_o^v \right\|^2 \right) + \left\| \sigma \boldsymbol{\Phi} \frac{\boldsymbol{\mathcal{I}}_o^l \overline{\boldsymbol{K}_{d_i}^l} + \boldsymbol{\mathcal{I}}_o^r \overline{\boldsymbol{K}_{d_i}^r}}{\left|\boldsymbol{K}_{d_i}^l\right|^2 + \left|\boldsymbol{K}_{d_i}^r\right|^2 + \sigma^2 |\boldsymbol{\Phi}|^2} \right\|^2 \right]. \end{aligned} \tag{D.15}$$

Then substituting $\mathcal{I}_o^v$ with Equation (D.5), we get:

$$
\begin{aligned}
&E\left(d_i|\boldsymbol{K}_{d^\star}^l,\boldsymbol{K}_{d^\star}^r,\sigma\right)\\
=&\mathbb{E}_{\boldsymbol{C}_i,\boldsymbol{\mathcal{N}}^l,\boldsymbol{\mathcal{N}}^r}\left[\left(\sum_{v=\{l,r\}}\left\|\boldsymbol{K}_{d_i}^v\frac{\left(\boldsymbol{K}_{d^\star}^l\boldsymbol{C}_i+\boldsymbol{\mathcal{N}}^l\right)\overline{\boldsymbol{K}_{d_i}^l}+\left(\boldsymbol{K}_{d^\star}^r\boldsymbol{C}_i+\boldsymbol{\mathcal{N}}^r\right)\overline{\boldsymbol{K}_{d_i}^r}}{\left|\boldsymbol{K}_{d_i}^l\right|^2+\left|\boldsymbol{K}_{d_i}^r\right|^2+\sigma^2\left|\boldsymbol{\Phi}\right|^2}-\left(\boldsymbol{K}_{d^\star}^v\boldsymbol{C}_i+\boldsymbol{\mathcal{N}}^v\right)\right\|^2\right)+\right.\\
&\left.\left\|\sigma\boldsymbol{\Phi}\frac{\left(\boldsymbol{K}_{d^\star}^l\boldsymbol{C}_i+\boldsymbol{\mathcal{N}}^l\right)\overline{\boldsymbol{K}_{d_i}^l}+\left(\boldsymbol{K}_{d^\star}^r\boldsymbol{C}_i+\boldsymbol{\mathcal{N}}^r\right)\overline{\boldsymbol{K}_{d_i}^r}}{\left|\boldsymbol{K}_{d_i}^l\right|^2+\left|\boldsymbol{K}_{d_i}^r\right|^2+\sigma^2\left|\boldsymbol{\Phi}\right|^2}\right\|^2\right].
\end{aligned}
\tag{D.16}
$$

We now define $B=\left|\boldsymbol{K}_{d_i}^l\right|^2+\left|\boldsymbol{K}_{d_i}^r\right|^2+\sigma^2\left|\boldsymbol{\Phi}\right|^2$. We can rearrange the above equation as:

$$
\begin{aligned}
&E\left(d_i|\boldsymbol{K}_{d^\star}^l,\boldsymbol{K}_{d^\star}^r,\sigma\right)\\
=&\mathbb{E}_{\boldsymbol{C}_i,\boldsymbol{\mathcal{N}}^l,\boldsymbol{\mathcal{N}}^r}\left[\left(\sum_{v=\{l,r\}}\left\|\frac{\boldsymbol{C}_i\left[\boldsymbol{K}_{d_i}^v\left(\boldsymbol{K}_{d^\star}^l\overline{\boldsymbol{K}_{d_i}^l}+\boldsymbol{K}_{d^\star}^r\overline{\boldsymbol{K}_{d_i}^r}\right)-\boldsymbol{K}_{d^\star}^vB\right]}{B}+\frac{\boldsymbol{K}_{d_i}^v\left(\boldsymbol{\mathcal{N}}^l\overline{\boldsymbol{K}_{d_i}^l}+\boldsymbol{\mathcal{N}}^r\overline{\boldsymbol{K}_{d_i}^r}\right)}{B}-\boldsymbol{\mathcal{N}}^v\right\|^2\right)+\right.\\
&\left.\left\|\sigma\boldsymbol{\Phi}\frac{\boldsymbol{C}_i\left(\boldsymbol{K}_{d^\star}^l\overline{\boldsymbol{K}_{d_i}^l}+\boldsymbol{K}_{d^\star}^r\overline{\boldsymbol{K}_{d_i}^r}\right)}{B}+\sigma\boldsymbol{\Phi}\frac{\boldsymbol{\mathcal{N}}^l\overline{\boldsymbol{K}_{d_i}^l}+\boldsymbol{\mathcal{N}}^r\overline{\boldsymbol{K}_{d_i}^r}}{B}\right\|^2\right].
\end{aligned}
\tag{D.17}
$$

Given that the entries of $\boldsymbol{\mathcal{N}}^{\{l,r\}}$ are independent identically distributed with distribution $\mathcal{N}\left(0,\sigma^2\right)$, we have $\mathbb{E}\left(\boldsymbol{\mathcal{N}}^v\right)=\boldsymbol{0},\mathbb{E}\left(\boldsymbol{\mathcal{N}}^{v2}\right)=\sigma^2$ and $\mathbb{E}\left(\boldsymbol{\mathcal{N}}^l\boldsymbol{\mathcal{N}}^r\right)=\boldsymbol{0}$, and we can simplify the above equation as:

$$
\begin{aligned}
&E\left(d_i|\boldsymbol{K}_{d^\star}^l,\boldsymbol{K}_{d^\star}^r,\sigma\right)\\
=&\mathbb{E}_{\boldsymbol{C}_i,\boldsymbol{\mathcal{N}}^l,\boldsymbol{\mathcal{N}}^r}\left[\left(\sum_{v=\{l,r\}}\left\|\frac{\boldsymbol{C}_i\left[\boldsymbol{K}_{d_i}^v\left(\boldsymbol{K}_{d^\star}^l\overline{\boldsymbol{K}_{d_i}^l}+\boldsymbol{K}_{d^\star}^r\overline{\boldsymbol{K}_{d_i}^r}\right)-\boldsymbol{K}_{d^\star}^vB\right]}{B}\right\|^2+\left\|\frac{\boldsymbol{K}_{d_i}^v\left(\boldsymbol{\mathcal{N}}^l\overline{\boldsymbol{K}_{d_i}^l}+\boldsymbol{\mathcal{N}}^r\overline{\boldsymbol{K}_{d_i}^r}\right)}{B}-\boldsymbol{\mathcal{N}}^v\right\|^2\right)+\right.\\
&\left.\left\|\sigma\boldsymbol{\Phi}\frac{\boldsymbol{C}_i\left(\boldsymbol{K}_{d^\star}^l\overline{\boldsymbol{K}_{d_i}^l}+\boldsymbol{K}_{d^\star}^r\overline{\boldsymbol{K}_{d_i}^r}\right)}{B}\right\|^2+\left\|\sigma\boldsymbol{\Phi}\frac{\boldsymbol{\mathcal{N}}^l\overline{\boldsymbol{K}_{d_i}^l}+\boldsymbol{\mathcal{N}}^r\overline{\boldsymbol{K}_{d_i}^r}}{B}\right\|^2\right]
\end{aligned}
\tag{D.18}
$$

$$
\begin{aligned}
=&\mathbb{E}_{\boldsymbol{C}_i}\left\{\left[\sum_{v=\{l,r\}}\left\|\frac{\boldsymbol{C}_i\left[\boldsymbol{K}_{d_i}^v\left(\boldsymbol{K}_{d^\star}^l\overline{\boldsymbol{K}_{d_i}^l}+\boldsymbol{K}_{d^\star}^r\overline{\boldsymbol{K}_{d_i}^r}\right)-\boldsymbol{K}_{d^\star}^vB\right]}{B}\right\|^2+\sigma^2\left(\left\|\frac{\boldsymbol{K}_{d_i}^{v\,2}+\sigma^2\left|\boldsymbol{\Phi}\right|^2}{B}\right\|^2+\left\|\frac{\boldsymbol{K}_{d_i}^l\boldsymbol{K}_{d_i}^r}{B}\right\|^2\right)\right]+\right.\\
&\left.\left\|\sigma\boldsymbol{\Phi}\frac{\boldsymbol{C}_i\left(\boldsymbol{K}_{d^\star}^l\overline{\boldsymbol{K}_{d_i}^l}+\boldsymbol{K}_{d^\star}^r\overline{\boldsymbol{K}_{d_i}^r}\right)}{B}\right\|^2+\sigma^2\left(\left\|\sigma\boldsymbol{\Phi}\frac{\boldsymbol{K}_{d_i}^l}{B}\right\|^2+\left\|\sigma\boldsymbol{\Phi}\frac{\boldsymbol{K}_{d_i}^r}{B}\right\|^2\right)\right\}.
\end{aligned}
\tag{D.19}
$$

Recall that, in Equation (D.10), we defined $\mathbf{\Phi}(f)$ such that $\frac{1}{|\mathbf{\Phi}(f)|^2} = \int_{\boldsymbol{C}_i} |\boldsymbol{C}_i(f)|^2 \mu(\boldsymbol{C}_i)$. Then we can further simplify $E\left(d_i | \boldsymbol{K}^l_{d^\star}, \boldsymbol{K}^r_{d^\star}, \sigma\right)$ as:

$$
\begin{aligned}
& E\left(d_i | \boldsymbol{K}^l_{d^\star}, \boldsymbol{K}^r_{d^\star}, \sigma\right) \\
=& \sum_f \left[\frac{\frac{1}{|\mathbf{\Phi}|^2} \left| \boldsymbol{K}^l_{d^\star} \boldsymbol{K}^r_{d_i} - \boldsymbol{K}^r_{d^\star} \boldsymbol{K}^l_{d_i} \right|^2}{B} \right] + \sum_f \left[\frac{\frac{1}{|\mathbf{\Phi}|^2} \sigma^2 |\mathbf{\Phi}|^2 \left(\left| \boldsymbol{K}^l_{d^\star} \right|^2 + \left| \boldsymbol{K}^r_{d^\star} \right|^2 \right)}{B} \right] + \\
& \sum_f \left[\sigma^2 \left(\left\| \frac{{\boldsymbol{K}^l_{d_i}}^2 + \sigma^2 |\mathbf{\Phi}|^2}{B} \right\|^2 + \left\| \frac{{\boldsymbol{K}^r_{d_i}}^2 + \sigma^2 |\mathbf{\Phi}|^2}{B} \right\|^2 + 2 \left\| \frac{\boldsymbol{K}^l_{d_i} \boldsymbol{K}^r_{d_i}}{B} \right\|^2 + \left\| \sigma \mathbf{\Phi} \frac{\boldsymbol{K}^l_{d_i}}{B} \right\|^2 + \left\| \sigma \mathbf{\Phi} \frac{\boldsymbol{K}^r_{d_i}}{B} \right\|^2 \right) \right] \qquad \text{(D.20)} \\
=& \sum_f \left[\frac{\frac{1}{|\mathbf{\Phi}|^2} \left| \boldsymbol{K}^l_{d^\star} \boldsymbol{K}^r_{d_i} - \boldsymbol{K}^r_{d^\star} \boldsymbol{K}^l_{d_i} \right|^2}{B} \right] + \sigma^2 \sum_f \left[\frac{\left| \boldsymbol{K}^l_{d^\star} \right|^2 + \left| \boldsymbol{K}^r_{d^\star} \right|^2}{B} + \frac{\sigma^2 |\mathbf{\Phi}|^2}{B} + 1 \right] \qquad \text{(D.21)} \\
=& \sum_f \left[\frac{\frac{1}{|\mathbf{\Phi}|^2} \left| \boldsymbol{K}^l_{d^\star} \boldsymbol{K}^r_{d_i} - \boldsymbol{K}^r_{d^\star} \boldsymbol{K}^l_{d_i} \right|^2}{\left| \boldsymbol{K}^l_{d_i} \right|^2 + \left| \boldsymbol{K}^r_{d_i} \right|^2 + \sigma^2 |\mathbf{\Phi}|^2} \right] + \sigma^2 \sum_f \left[\frac{\left| \boldsymbol{K}^l_{d^\star} \right|^2 + \left| \boldsymbol{K}^r_{d^\star} \right|^2 + \sigma^2 |\mathbf{\Phi}|^2}{\left| \boldsymbol{K}^l_{d_i} \right|^2 + \left| \boldsymbol{K}^r_{d_i} \right|^2 + \sigma^2 |\mathbf{\Phi}|^2} + 1 \right]. \qquad \text{(D.22)}
\end{aligned}
$$

If we define $C_1\left(\boldsymbol{K}^{\{l,r\}}_{d_i}, \sigma, \mathbf{\Phi}\right) = \frac{\frac{1}{|\mathbf{\Phi}|^2}}{\left|\boldsymbol{K}^l_{d_i}\right|^2 + \left|\boldsymbol{K}^r_{d_i}\right|^2 + \sigma^2 |\mathbf{\Phi}|^2}$, and $C_2(\sigma) = \sigma^2 \sum_f 1$, then Equation (D.22) boils down to Equation (5.5).

D.2 Edge-aware Total Variation Function

We first define a pixel-wise total variation function of a single-layer image $\mathbf{I}$ that in used in both the intensity smoothness prior $\mathcal{L}_{\text{intensity}}$ and the alpha and transmittance smoothness prior $\mathcal{L}_{\text{alpha}}$:

$$
\boldsymbol{V}(\mathbf{I}) = \sqrt{\mathbf{I}^2 * g - (\mathbf{I} * g)^2}, \qquad \text{(D.23)}
$$

where g is a two-dimensional Gaussian blur kernel:

$$
g = \begin{bmatrix} 1/16 & 1/8 & 1/16 \\ 1/8 & 1/4 & 1/8 \\ 1/16 & 1/8 & 1/16 \end{bmatrix}. \qquad \text{(D.24)}
$$

Each "pixel" in $\boldsymbol{V}(\mathbf{I})(x, y)$ is equivalent to, for the 3×3 window surrounding pixel (x, y) in $\mathbf{I}$, computing the sample standard deviation (weighted by a Gaussian kernel) of the pixel intensities in that window. This follows easily from two facts: 1) as g sums to 1 by construction, $\mathbf{I} * g$ produces an image whose pixel intensities can be viewed as expectations of their surrounding 3×3 input patch; and 2) the standard deviation $\sqrt{\mathbb{E}\left[(X - \mathbb{E}[X])^2\right]}$ can be written equivalently as $\sqrt{\mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2}$.

As done in prior work [177], we would like to encourage edge-aware smoothness in addition to minimizing total variation, so a bilateral edge mask is computed using this total variation:

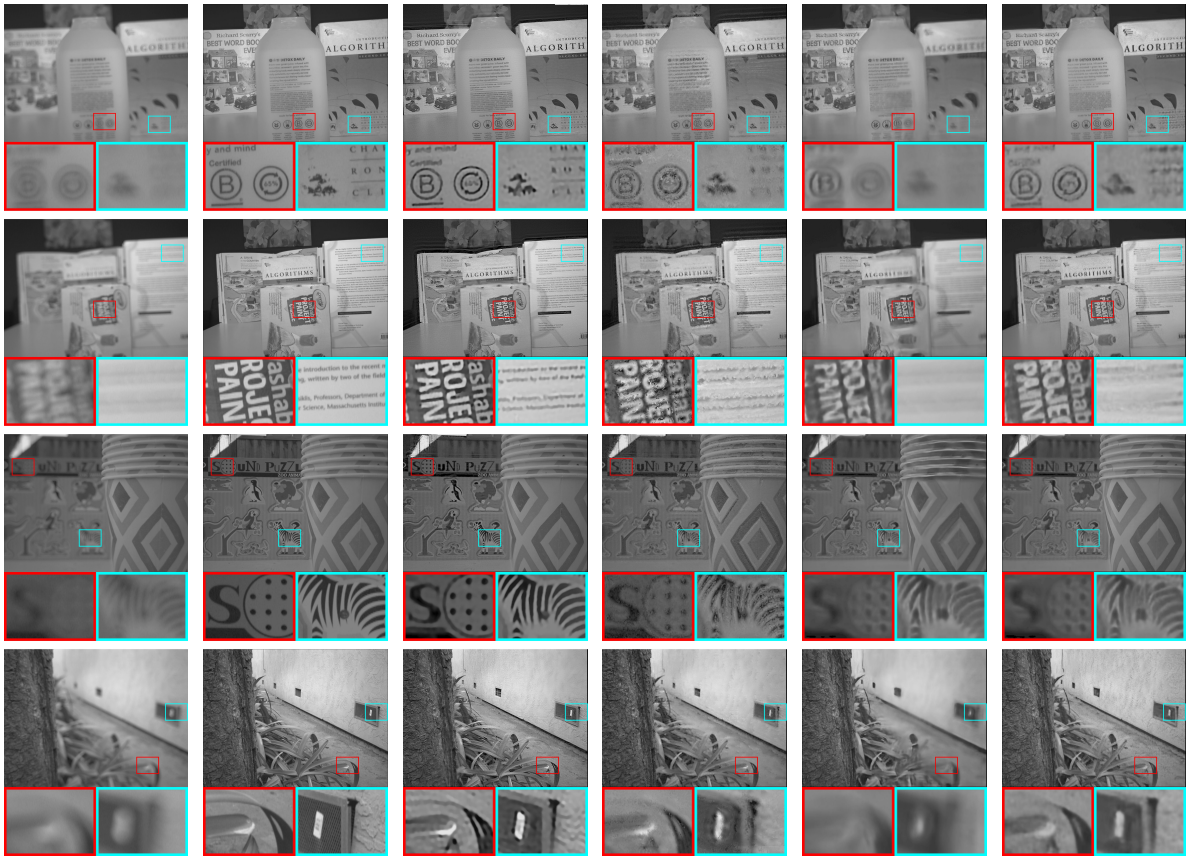
$$\boldsymbol{E}(\mathbf{I}) = \mathbf{1} - \exp\left(-\frac{\mathbf{I}^2 * g - (\mathbf{I} * g)^2}{2\beta^2}\right). \tag{D.25}$$

In this equation, β is set to $1/32$ (assuming pixel intensities are in $[0, 1]$). A joint total variation function that takes into account both the original and the edge-aware total variation is then defined as:

$$\boldsymbol{V}_{\boldsymbol{E}}(\mathbf{I}, \boldsymbol{E}) = \ell\left(\boldsymbol{V}(\mathbf{I})\right) + (1 - \boldsymbol{E}) \odot \ell\left(\boldsymbol{V}(\mathbf{I})\right). \tag{D.26}$$

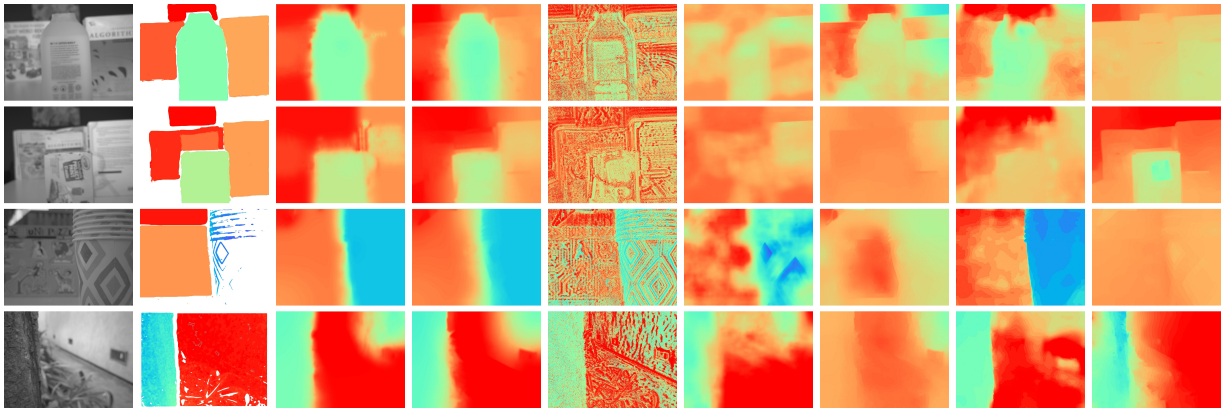
D.3 Additional Experimental Results

We show qualitative results on more data in our collected Google Pixel 4 dataset in Figures D.1-D.3.



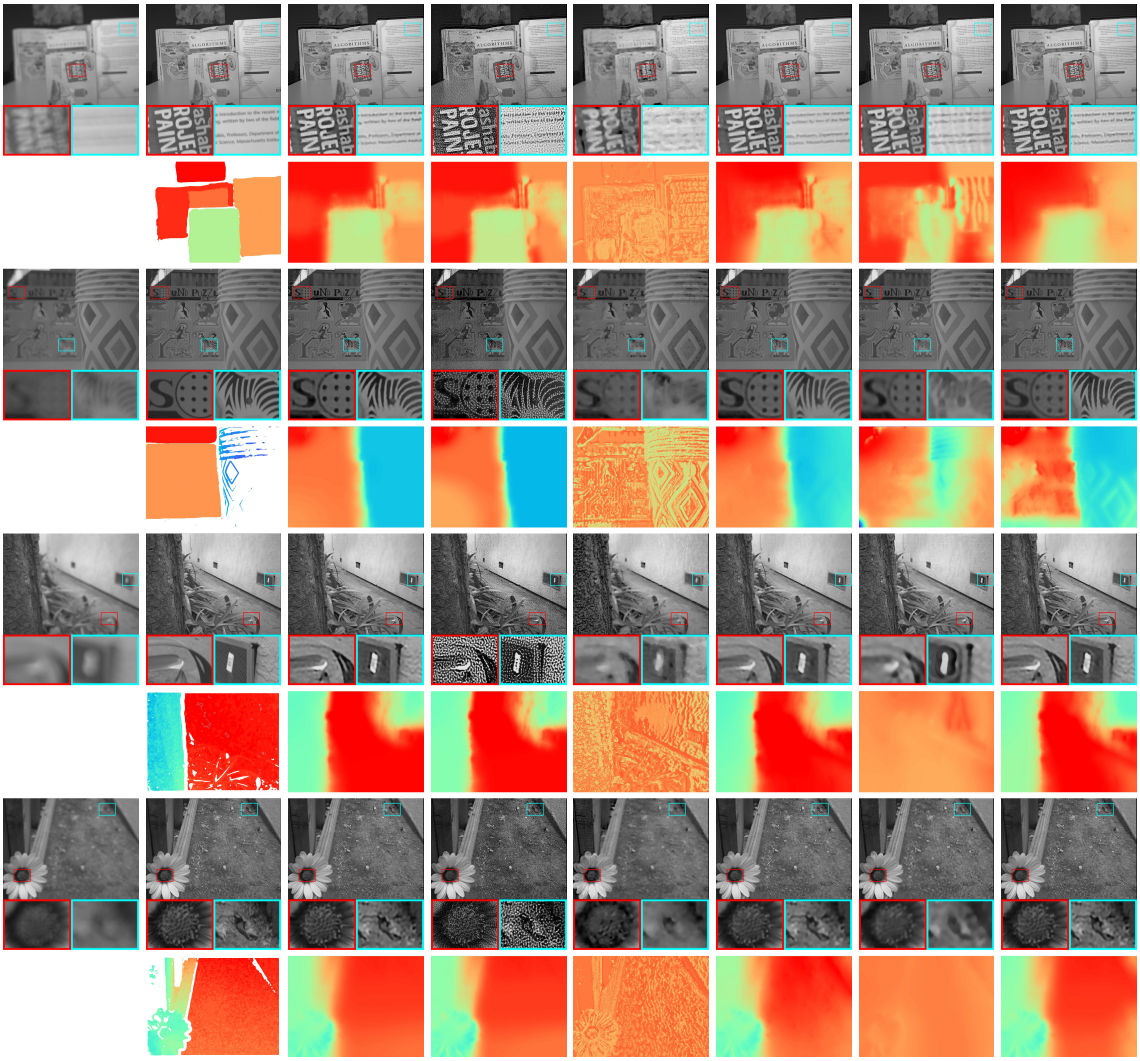
(a) input image (b) GT all-in-focus image (c) ours (d) Wiener deconv. [194] (e) DPDNet [1] (f) DPDNet w/ calib [1]

Figure D.1: More qualitative comparisons of various defocus deblurring methods.



(a) input image (b) ground truth (c) ours (d) ours w/ GF (e) Wiener [194] (f) DMENet [100] (g) [144] (h) Garg [53] (i) Wadhwa [181]

Figure D.2: More qualitative comparisons of defocus map estimation methods.



(a) Input image (b) Ground truth (c) Ours full (d) No $\mathcal{L}_{\text{intensity}}$ (e) No $\mathcal{L}_{\text{alpha}}$ (f) No $\mathcal{L}_{\text{entropy}}$ (g) No $\mathcal{L}_{\text{aux}}$ (h) No $\mathcal{B}$

Figure D.3: More qualitative results on ablation study.

Bibliography

[1] Abdullah Abuolaim and Michael S. Brown. Defocus deblurring using dual-pixel data. *European Conference on Computer Vision*, 2020. (document), 5.1, 5.4.1, 5.4.3, 5.4.4, 5.8(e), 5.8(f), **??**, 5.4.4, 5.12(a), 5.12(b), 5.12, D.1(e), D.1(f)

[2] Abdullah Abuolaim, Abhijith Punnappurath, and Michael S. Brown. Revisiting autofocus for smartphone cameras. *European Conference on Computer Vision*, 2018. 1.3.4, 5

[3] Amit Agrawal, Yuichi Taguchi, and Srikumar Ramalingam. Analytical forward projection for axial non-central dioptric and catadioptric cameras. In *European Conference on Computer Vision*, pages 129–143. Springer, 2010. A.5

[4] Amit Agrawal, Yuichi Taguchi, and Srikumar Ramalingam. Beyond alhazen's problem: Analytical projection model for non-central catadioptric cameras with quadric mirrors. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 2993–3000. IEEE, 2011. A.5

[5] Byeongjoo Ahn, Akshat Dave, Ashok Veeraraghavan, Ioannis Gkioulekas, and Aswin C Sankaranarayanan. Convolutional approximations to the general non-line-of-sight imaging operator. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7889–7899, 2019. 6.1

[6] Emma Alexander, Qi Guo, Sanjeev Koppal, Steven Gortler, and Todd Zickler. Focal flow: Measuring distance and velocity with defocus and differential motion. In *European conference on computer vision*, pages 667–682. Springer, 2016. 1.1

[7] Emma Alexander, Qi Guo, Sanjeev Koppal, Steven J Gortler, and Todd Zickler. Focal flow: Velocity and depth from differential defocus through motion. *International Journal of Computer Vision*, 126(10):1062–1083, 2018. 1.1

[8] Sameer Ansari, Neal Wadhwa, Rahul Garg, and Jiawen Chen. Wireless software

synchronization of multiple distributed cameras. *IEEE International Conference on Computational Photography*, 2019. 5.1

[9] Victor Arellano, Diego Gutierrez, and Adrian Jarabo. Fast back-projection for non-line of sight reconstruction. *Optics Express*, 25(10):11574–11583, 2017. 3.1

[10] Amir Atapour-Abarghouei and Toby P. Breckon. Real-time monocular depth estimation using synthetic data with domain adaptation via image style transfer. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018. 5.1

[11] Ruzena Bajcsy and Lawrence Lieberman. Texture gradient as a depth cue. *Computer Graphics and Image Processing*, 1976. 5.1

[12] Simon Baker and Shree K Nayar. A theory of single-viewpoint catadioptric image formation. *International Journal of Computer Vision*, 35(2):175–196, 1999. A.5

[13] Sanjay Bakshi and Yee-Hong Yang. Shape from shading for non-lambertian surfaces. In *Proceedings of 1st International Conference on Image Processing*, volume 2, pages 130–134. IEEE, 1994. 4

[14] Jonathan Balzer, Stefan Werling, and Jürgen Beyerer. Regularization of the deflectometry problem using shading data. In *Two-and Three-Dimensional Methods for Inspection and Metrology IV*, volume 6382, pages 66–76. SPIE, 2006. 4

[15] Manel Baradad, Vickie Ye, Adam B Yedidia, Frédo Durand, William T Freeman, Gregory W Wornell, and Antonio Torralba. Inferring light fields from shadows. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6267–6275, 2018. 3.1

[16] John L Barron, David J Fleet, and Steven S Beauchemin. Performance of optical flow techniques. *International journal of computer vision*, 12(1):43–77, 1994. 1.2.1

[17] Jonathan T. Barron. A general and adaptive robust loss function. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 5.3.3

[18] Jonathan T. Barron and Jitendra Malik. Shape, albedo, and illumination from a single image of an unknown object. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2012. 5.1

[19] Ronen Basri, David Jacobs, and Ira Kemelmacher. Photometric stereo with general, unknown lighting. *International Journal of computer vision*, 72(3):239–257, 2007. 1.3.3

[20] Mufeed Batarseh, Sergey V Sukhov, Zhiqin Shen, Heath Gemar, Reza Rezvani, and Aristide Dogariu. Passive sensing around the corner using spatial coherence. *Nature communications*, 9(1):3629, 2018. 3.1

[21] Jacopo Bertolotti, Elbert G van Putten, Christian Blum, Ad Lagendijk, Willem L Vos, and Allard P Mosk. Non-invasive imaging through opaque scattering layers. *Nature*, 491(7423):232, 2012. 3.1

[22] Andrew Blake and G Brelstaff. Specular stereo. In *IJCAI*, volume 2, pages 973–976. Citeseer, 1985. 1.3.3

[23] Jeremy Boger-Lombard and Ori Katz. Non line-of-sight localization by passive optical time-of-flight. *ArXiv e-prints*, August 2018. 3.1

[24] Katherine L. Bouman, Vickie Ye, Adam B. Yedidia, Frédo Durand, Gregory W Wornell, Antonio Torralba, and William T Freeman. Turning corners into cameras: Principles and methods. In *ICCV*, 2017. 3.1

[25] Michael Brady and Alan Yuille. An extremum principle for shape from contour. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1984. 5.1

[26] James W Bruce and Peter J Giblin. *Curves and Singularities: a geometrical introduction to singularity theory*. Cambridge university press, 1992. 2.5.1, A.7, A.7, A.7

[27] James W Bruce, Peter J Giblin, and Christopher G Gibson. On caustics of plane curves. *The American Mathematical Monthly*, 88(9):651–667, 1981. 2.5.1, A.7, A.7, A.7

[28] Jan Burke, Alexey Pak, Sebastian Höfer, Mathias Ziebarth, Masoud Roschani, and Jürgen Beyerer. Deflectometry for specular surfaces: an overview. *arXiv preprint arXiv:2204.11592*, 2022. 1.3.3, 4.2

[29] Mauro Buttafava, Jessica Zeman, Alberto Tosi, Kevin Eliceiri, and Andreas Velten. Non-line-of-sight imaging using a time-gated single photon avalanche diode. *Optics express*, 23(16):20997–21011, 2015. 3.1, 3.2

[30] Manmohan Chandraker. What camera motion reveals about shape with unknown brdf. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2171–2178, 2014. 1.2.1

[31] Manmohan Chandraker. On shape and material recovery from motion. In *European Conference on Computer Vision*, pages 202–217. Springer, 2014. 1.2.1

[32] Manmohan Chandraker. The information available to a moving observer on shape with unknown, isotropic brdfs. *IEEE transactions on pattern analysis and machine intelligence*, 38(7):1283–1297, 2015. 1.2.1

[33] Manmohan Chandraker, Jiamin Bai, and Ravi Ramamoorthi. A theory of photometric reconstruction for unknown isotropic reflectances. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, 2011. (document), 1.3, 1.1, 1.2.1

[34] Manmohan Chandraker, Jiamin Bai, and Ravi Ramamoorthi. On differential photometric reconstruction for unknown, isotropic brdfs. *IEEE transactions on pattern analysis and machine intelligence*, 35(12):2941–2955, 2012. 1.2.1

[35] Manmohan Chandraker, Dikpal Reddy, Yizhou Wang, and Ravi Ramamoorthi. What object motion reveals about shape with unknown brdf and lighting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2523–2530, 2013. 1.2.1

[36] Pierre Charbonnier, Laure Blanc-Feraud, Gilles Aubert, and Michel Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. *IEEE International Conference on Image Processing*, 1994. 5.3.3

[37] Ching-Hui Chen, Hui Zhou, and Timo Ahonen. Blur-aware disparity estimation from defocus stereo images. *IEEE/CVF International Conference on Computer Vision*, 2015. 5.1

[38] Min Chen and James Arvo. Perturbation methods for interactive specular reflections. *IEEE Transactions on Visualization and Computer Graphics*, 6(3):253–264, 2000. 2.3

[39] Min Chen and James Arvo. Theory and application of specular path perturbation. *ACM Transactions on Graphics (TOG)*, 19(4):246–278, 2000. 2.4.1, 2.4.1

[40] Sunghwan Choi, Dongbo Min, Bumsub Ham, Youngjung Kim, Changjae Oh, and Kwanghoon Sohn. Depth analogy: Data-driven approach for single image depth estimation using gradient samples. *IEEE Transactions on Image Processing*, 2015. 5.1

[41] James J Clark. Active photometric stereo. In *CVPR*, volume 92, pages 29–34, 1992. 1.3.3

[42] Laurent D'Andrès, Jordi Salvador, Axel Kochale, and Sabine Süsstrunk. Non-parametric blur map regression for depth of field extension. *IEEE Transactions on Image Processing*, 2016. 5.1

[43] Edward R. Dowski and W. Thomas Cathey. Extended depth of field through wave-front coding. *Applied optics*, 1995. 1.3.4, 5

[44] Dual Pixel Capture App. Dual pixel capture app. https://github.com/google-research/google-research/tree/master/dual_pixels. 5.4.1

[45] Philip Dutré, Kavita Bala, and Philippe Bekaert. *Advanced global illumination*. AK Peters, Ltd., 2006. 3.2

[46] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network. *Advances in Neural Information Processing Systems*, 2014. 5.1

[47] Jose M. Facil, Benjamin Ummenhofer, Huizhong Zhou, Luis Montesano, Thomas Brox, and Javier Civera. Cam-convs: camera-aware multi-scale convolutions for single-view depth. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 5.1

[48] Paolo Favaro. Recovering thin structures via nonlocal-means regularization with application to depth from defocus. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2010. 1.1, 1.3.4, 5.1

[49] Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T. Roweis, and William T. Freeman. Removing camera shake from a single photograph. *ACM Transactions on Graphics*, 2006. 5.5

[50] Steven H Ferris. Motion parallax and absolute distance. *Journal of experimental psychology*, 95(2):258, 1972. 1.1

[51] D.A. Fish, A.M. Brinicombe, E.R. Pike, and J.G. Walker. Blind deconvolution by means of the Richardson–Lucy algorithm. *Journal of the Optical Society of America A*, 1995. 5.1

[52] Huan Fu, Mingming Gong, Chaohui Wang, Kayhan Batmanghelich, and Dacheng Tao. Deep ordinal regression network for monocular depth estimation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018. 5.1

[53] Rahul Garg, Neal Wadhwa, Sameer Ansari, and Jonathan T. Barron. Learning sin-

gle camera depth estimation using dual-pixels. *IEEE/CVF International Conference on Computer Vision*, 2019. (document), 5.1, 5.2, 5.4.1, 5.4.4, 5.9(h), **??**, 5.1, D.2(h)

[54] Ravi Garg, Vijay Kumar B.G., Gustavo Carneiro, and Ian Reid. Unsupervised cnn for single view depth estimation: Geometry to the rescue. *European Conference on Computer Vision*, 2016. 5.1

[55] Genevieve Gariepy, Nikola Krstajić, Robert Henderson, Chunyong Li, Robert R Thomson, Gerald S Buller, Barmak Heshmat, Ramesh Raskar, Jonathan Leach, and Daniele Faccio. Single-photon sensitive light-in-fight imaging. *Nature communications*, 6:6021, 2015. 3.1, 3.4.1

[56] Ioannis Gkioulekas, Anat Levin, Frédo Durand, and Todd Zickler. Micron-scale light transport decomposition using interferometry. *ACM Transactions on Graphics*, 2015. 3.1, 3.4.2

[57] Clément Godard, Oisin Mac Aodha, and Gabriel J. Brostow. Unsupervised monocular depth estimation with left-right consistency. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2017. 5.1

[58] Clément Godard, Oisin Mac Aodha, and Gabriel J. Brostow. Digging into self-supervised monocular depth estimation. *IEEE/CVF International Conference on Computer Vision*, 2019. 5.1

[59] Paul Grossmann. Depth from focus. *Pattern Recognition Letters*, 1987. 5.1

[60] Qi Guo, Zhujun Shi, Yao-Wei Huang, Emma Alexander, Cheng-Wei Qiu, Federico Capasso, and Todd Zickler. Compact single-shot metalens depth sensors inspired by eyes of jumping spiders. *Proceedings of the National Academy of Sciences*, 116(46): 22959–22965, 2019. (document), 1.4, 1.1

[61] Xiaoyang Guo, Hongsheng Li, Shuai Yi, Jimmy Ren, and Xiaogang Wang. Learning monocular depth by distilling cross-domain stereo networks. *European Conference on Computer Vision*, 2018. 5.1

[62] Otkrist Gupta, Thomas Willwacher, Andreas Velten, Ashok Veeraraghavan, and Ramesh Raskar. Reconstruction of hidden 3d shapes using diffuse reflections. *Optics express*, 20(17):19096–19108, 2012. 3.1

[63] Christian Häne, L'ubor Ladický, and Marc Pollefeys. Direction matters: Depth estimation with a surface normal classifier. *IEEE/CVF Conference on Computer Vision*

and Pattern Recognition, 2015. 5.1

[64] Richard Hartley. A. zisserman multiple view geometry in computer vision, 2000. 1.3.3

[65] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003. 5.1

[66] Samuel W. Hasinoff, Frédo Durand, and William T. Freeman. Noise-optimal capture for high dynamic range photography. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2010. 5.3.2

[67] Samuel W. Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T. Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics*, 2016. 5.3.2

[68] Caner Hazırbaş, Sebastian Georg Soyer, Maximilian Christian Staab, Laura Leal-Taixé, and Daniel Cremers. Deep depth from focus. *Asian Conference on Computer Vision*, 2018. 5.1

[69] Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. *European Conference on Computer Vision*, 2010. 5.4.4

[70] Felix Heide, Lei Xiao, Wolfgang Heidrich, and Matthias B Hullin. Diffuse mirrors: 3d reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3222–3229, 2014. 3.1

[71] Felix Heide, Matthew O'Toole, Kai Zhang, David Lindell, Steven Diamond, and Gordon Wetzstein. Robust non-line-of-sight imaging with single photon detectors. *arXiv preprint arXiv:1711.07134*, 2017. 3.1

[72] Helicon Focus. Helicon focus. https://www.heliconsoft.com/. 5.4.1

[73] Quercus Hernandez, Diego Gutierrez, and Adrian Jarabo. A computational model of a single-photon avalanche diode sensor for transient imaging. *arXiv preprint arXiv:1703.02635*, 2017. 3.1, 3.2.2, 3.4.3

[74] Charles Herrmann, Richard Strong Bowen, Neal Wadhwa, Rahul Garg, Qiurui He, Jonathan T. Barron, and Ramin Zabih. Learning to autofocus. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 1.3.4, 5

[75] Derek Hoiem, Alexei A. Efros, and Martial Hebert. Automatic photo pop-up. *ACM Transactions on Graphics*, 2005. 5.1

[76] Berthold K.P. Horn. Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. Technical report, Massachusetts Institute of Technology, 1970. 2.3.1, 5.1

[77] Berthold KP Horn and Brian G Schunck. Determining optical flow. *Artificial intelligence*, 17(1-3):185–203, 1981. 1.2.1

[78] Ivo Ihrke, Kiriakos N Kutulakos, Hendrik PA Lensch, Marcus Magnor, and Wolfgang Heidrich. Transparent and specular object reconstruction. In *Computer Graphics Forum*, volume 29, pages 2400–2426. Wiley Online Library, 2010. 1.3.3

[79] Katsushi Ikeuchi. Determining surface orientations of specular surfaces by using the photometric stereo method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):661–669, 1981. 4

[80] Wenzel Jakob. Mitsuba renderer, 2010. http://www.mitsuba-renderer.org. (document), 3.4.3, 4, 4.3(b), 4.3, 4.4.1

[81] Wenzel Jakob and Steve Marschner. Manifold exploration: A markov chain monte carlo technique for rendering scenes with difficult specular transport. *ACM Transactions on Graphics*, 2012. 2.3, 2.4.1, A.5

[82] Adrian Jarabo, Julio Marco, Adolfo Muñoz, Raul Buisan, Wojciech Jarosz, and Diego Gutierrez. A Framework for Transient Rendering. *ACM Trans. Graph.*, 33 (6):177:1–177:10, November 2014. ISSN 0730-0301. 3.1

[83] Adrian Jarabo, Belen Masia, Julio Marco, and Diego Gutierrez. Recent advances in transient imaging: A computer graphics and vision perspective. *Visual Informatics*, 1(1):65–79, 2017. 3.2

[84] Huaizu Jiang, Erik G. Learned-Miller, Gustav Larsson, Michael Maire, and Greg Shakhnarovich. Self-supervised depth learning for urban scene understanding. *European Conference on Computer Vision*, 2018. 5.1

[85] Achuta Kadambi, Hang Zhao, Boxin Shi, and Ramesh Raskar. Occluded imaging with time-of-flight sensors. *ACM Transactions on Graphics (ToG)*, 35(2):15, 2016. 3.1

[86] Ali Karaali and Claudio Rosito Jung. Edge-based defocus blur estimation with adaptive scale selection. *IEEE Transactions on Image Processing*, 2017. 5.1

[87] Ori Katz, Eran Small, and Yaron Silberberg. Looking around corners and through thin turbid layers in real time with scattered incoherent light. *Nature photonics*, 6 (8):549–553, 2012. 3.1

[88] Ori Katz, Pierre Heidmann, Mathias Fink, and Sylvain Gigan. Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations. *Nature photonics*, 8(10):784, 2014. 3.1

[89] Michael Kazhdan, Matthew Bolitho, and Hugues Hoppe. Poisson surface reconstruction. In *Proc. Eurographics Symposium on Geometry Processing*, pages 61–70, 2006. 2.4.1, 2.4.1

[90] Diederick P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 2015. 5.4.3

[91] Ahmed Kirmani, Tyler Hutchison, James Davis, and Ramesh Raskar. Looking around the corner using transient imaging. In *2009 IEEE 12th International Conference on Computer Vision*, pages 159–166. IEEE, 2009. 1.3.2

[92] Ahmed Kirmani, Tyler Hutchison, James Davis, and Ramesh Raskar. Looking around the corner using ultrafast transient imaging. *International journal of computer vision*, 95(1):13–28, 2011. 3.1

[93] Jonathan Klein, Christoph Peters, Jaime Martín, Martin Laurenzis, and Matthias B Hullin. Tracking objects outside the line of sight using 2d intensity images. *Scientific reports*, 6, 2016. 3.1

[94] Markus C Knauer, Jurgen Kaminski, and Gerd Hausler. Phase measuring deflectometry: a new approach to measure specular free-form surfaces. In *Optical Metrology in Production Engineering*, volume 5457, pages 366–376. SPIE, 2004. 1.3.3

[95] Janusz Konrad, Meng Wang, Prakash Ishwar, Chen Wu, and Debargha Mukherjee. Learning-based, automatic 2d-to-3d image and video conversion. *IEEE Transactions on Image Processing*, 2013. 5.1

[96] Dilip Krishnan and Rob Fergus. Fast image deconvolution using hyper-laplacian priors. *Advances in Neural Information Processing Systems*, 2009. 5.1

[97] Kiriakos N Kutulakos and Eron Steger. A theory of refractive and specular 3d shape by light-path triangulation. *International Journal of Computer Vision*, 76(1): 13–29, 2008. 1.3.3

[98] Marco La Manna, Fiona Kine, Eric Breitbach, Jonathan Jackson, and Andreas Velten. Error backprojection algorithms for non-line-of-sight imaging. 3.1

[99] Lubor Ladicky, Jianbo Shi, and Marc Pollefeys. Pulling things out of perspective. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2014. 5.1

[100] Junyong Lee, Sungkil Lee, Sunghyun Cho, and Seungyong Lee. Deep defocus map estimation using domain adaptation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2019. 5.1, 5.4.4, 5.9(f), **??**, D.2(f)

[101] Anat Levin, Rob Fergus, Frédo Durand, and William T. Freeman. Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics*, 2007. 1.3.4, 5, 5.1, 5.3.3

[102] Anat Levin, Samuel W. Hasinoff, Paul Green, Frédo Durand, and William T. Freeman. 4D frequency analysis of computational cameras for depth of field extension. *ACM Transactions on Graphics*, 2009. 1.3.4, 5

[103] Anat Levin, Yair Weiss, Frédo Durand, and William T. Freeman. Understanding blind deconvolution algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011. 5.5

[104] Marc Levoy, Ren Ng, Andrew Adams, Matthew Footer, and Mark Horowitz. Light field microscopy. In *ACM SIGGRAPH 2006 Papers*, pages 924–934. 2006. 1.1

[105] Feng Li, Jian Sun, Jue Wang, and Jingyi Yu. Dual-focus stereo imaging. *Journal of Electronic Imaging*, 2010. 5.1

[106] Jun Li, Reinhard Klein, and Angela Yao. A two-streamed network for estimating fine-scaled depth maps from single rgb images. *IEEE/CVF International Conference on Computer Vision*, 2017. 5.1

[107] Xiu Li, Hongwei Qin, Yangang Wang, Yongbing Zhang, and Qionghai Dai. DEPT: depth estimation by parameter transfer for single still images. *Asian Conference on Computer Vision*, 2014. 5.1

[108] David Lindell, Gordon Wetzstein, and Vladlen Koltun. Acoustic non-line-of-sight imaging. In *CVPR*, 2019. 3.5

[109] Hugh W Lippincott and Henry Stark. Optical–digital detection of dents and scratches on specular metal surfaces. *Applied optics*, 21(16):2875–2881, 1982. 1.3.3

[110] Chao Liu, Srinivasa G Narasimhan, and Artur W Dubrawski. Near-light photometric stereo using circularly placed point light sources. In *2018 IEEE International Conference on Computational Photography (ICCP)*, pages 1–10. IEEE, 2018. (document), 1.3, 1.1

[111] Fayao Liu, Chunhua Shen, Guosheng Lin, and Ian Reid. Learning depth from single monocular images using deep convolutional neural fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015. 5.1

[112] Xiaochun Liu, Sebastian Bauer, and Andreas Velten. Analysis of feature visibility in non-line-of-sight measurements. In *CVPR*, 2019. 3.5

[113] Xiaochun Liu, Ibón Guillén, Marco La Manna, Ji Hyun Nam, Syed Azer Reza, Toan Huu Le, Adrian Jarabo, Diego Gutierrez, and Andreas Velten. Non-line-of-sight imaging using phasor-field virtual wave optics. *Nature*, 2019. 6.1

[114] Patric Ljung, Jens Krüger, Eduard Groller, Markus Hadwiger, Charles D. Hansen, and Anders Ynnerman. State of the art in transfer functions for direct volume rendering. *Computer Graphics Forum*, 2016. 5.3.1

[115] Bruce D Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. 1981. 1.2.1

[116] Reza Mahjourian, Martin Wicke, and Anelia Angelova. Unsupervised learning of depth and ego-motion from monocular video using 3D geometric constraints. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018. 5.1

[117] Fahim Mannan and Michael S. Langer. Blur calibration for depth from defocus. *Conference on Computer and Robot Vision*, 2016. 5.2, 5.4.2

[118] Nikolaus Mayer, Eddy Ilg, Philipp Fischer, Caner Hazirbas, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. What makes good synthetic training data for learning disparity and optical flow estimation? *International Journal of Computer Vision*, 2018. 5.1

[119] Lingfei Meng, Liyang Lu, Noah Bedard, and Kathrin Berkner. Single-shot specular surface reconstruction with gonio-plenoptic imaging. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3433–3441, 2015. 1.3.3, 4

[120] Tomer Michaeli and Michal Irani. Blind deblurring using internal patch recurrence. *European Conference on Computer Vision*, 2014. 5.1

[121] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *European Conference on Computer Vision*, 2020. 5.5

[122] Don Mitchell and Pat Hanrahan. Illumination from curved reflectors. In *ACM SIGGRAPH Computer Graphics*, volume 26, pages 283–291. ACM, 1992. 2.3, 2.4.1, 4.3.1, 4.3.1, 4.3.1, 4.3.3

[123] Nigel JW Morris and Kiriakos N Kutulakos. Dynamic refraction stereo. *IEEE transactions on pattern analysis and machine intelligence*, 33(8):1518–1531, 2011. 6.1

[124] John C Mott-Smith et al. *Medial axis transformations*, volume 212. Academic, 1970. 2.2

[125] Hajime Nagahara, Sujit Kuthirummal, Changyin Zhou, and Shree K. Nayar. Flexible Depth of Field Photography. *European Conference on Computer Vision*, 2008. 5.1

[126] Jogendra Nath Kundu, Phani Krishna Uppala, Anuj Pahuja, and R. Venkatesh Babu. Adadepth: Unsupervised content congruent adaptation for depth estimation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018. 5.1

[127] Elizabeth Nawrot and Mark Nawrot. The role of eye movements in depth from motion parallax during infancy. *Journal of vision*, 13(14):15–15, 2013. 1.1

[128] Shree K Nayar, Arthur C Sanderson, Lee E Weiss, and David A Simon. Specular surface inspection using structured highlight and gaussian images. *IEEE Transactions on Robotics and Automation*, 6(2):208–218, 1990. 1.3.3

[129] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan. Light field photography with a hand-held plenoptic camera. Technical report, Stanford University, 2005. 1.3.4, 5

[130] Michael Oren and Shree K Nayar. A theory of specular surface geometry. *International Journal of Computer Vision*, 24(2):105–124, 1997. 1.3.3

[131] Stanley Osher and Ronald Fedkiw. *Level set methods and dynamic implicit surfaces*, volume 153. Springer Science & Business Media, 2006. 2.2.1

[132] M. O'Toole, R. Raskar, and K.N. Kutulakos. Primal-dual coding to probe light transport. *ACM Transactions on Graphics*, 2012. 3.4.2

[133] Matthew O'Toole, Felix Heide, David B Lindell, Kai Zang, Steven Diamond, and Gordon Wetzstein. Reconstructing transient images from single-photon sensors. *CVPR*, 2017. 3.1, 3.4.1

[134] Matthew O'Toole, Felix Heide, Kai Lindell, David B.and Zang, Steven Diamond, and Gordon Wetzstein. Reconstructing transient images from single-photon sensors. In *CVPR*, 2017. 3.1

[135] Matthew O'Toole, David B Lindell, and Gordon Wetzstein. Confocal non-line-of-sight imaging based on the light-cone transform. *Nature*, 555(7696):338, 2018. 3.1, 3.2, 3.4.1, 3.5

[136] Matthew O'Toole, David B Lindell, and Gordon Wetzstein. Confocal non-line-of-sight imaging based on the light-cone transform. *Nature*, 555(7696):338–341, 2018. 6.1

[137] Jinsun Park, Yu-Wing Tai, Donghyeon Cho, and In So Kweon. A unified approach of multi-scale deep and hand-crafted features for defocus estimation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2017. 5.1

[138] Adithya Pediredla, Mauro Buttafava, Alberto Tosi, Oliver Cossairt, and Ashok Veeraraghavan. Reconstructing rooms using photon echoes: A plane based model and reconstruction algorithm for looking around the corner. *ICCP*, 2017. 3.1, 3.2

[139] Adithya Pediredla, Akshat Dave, and Ashok Veeraraghavan. Snlos: Non-line-of-sight scanning through temporal focusing. *ICCP*, 2019. 3.1

[140] Adithya Pediredla, Ashok Veeraraghavan, and Ioannis Gkioulekas. Elliptic path sampling for time-gated rendering. *ACM Transactions on Graphics (TOG)*, 2019. 3.2

[141] Alex Paul Pentland. A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1987. 1.1, 1.3.4, 5.1

[142] Tsai Ping-Sing and Mubarak Shah. Shape from shading using linear approximation. *Image and Vision computing*, 12(8):487–498, 1994. 4

[143] Abhijith Punnappurath and Michael S. Brown. Reflection removal using a dual-pixel sensor. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 5.1

[144] Abhijith Punnappurath, Abdullah Abuolaim, Mahmoud Afifi, and Michael S. Brown. Modeling defocus-disparity in dual-pixel sensors. *IEEE International Con-*

ference on Computational Photography, 2020. (document), 5.1, 5.2, 5.3(e), 5.3, 5.3, 5.4.1, 5.4.4, 5.4.4, 5.9(g), **??**, D.2(g)

[145] Yvain Quéau, Jean-Denis Durou, and Jean-François Aujol. Normal integration: a survey. *Journal of Mathematical Imaging and Vision*, 60(4):576–593, 2018. 2.3.1

[146] Hossein Ragheb and Edwin R Hancock. A probabilistic framework for specular shape-from-shading. *Pattern Recognition*, 36(2):407–427, 2003. 4

[147] Rene Ranftl, Vibhav Vineet, Qifeng Chen, and Vladlen Koltun. Dense monocular depth estimation in complex dynamic scenes. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2016. 5.1

[148] René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 5.1

[149] Alfréd Rényi. On measures of entropy and information. *Berkeley Symposium on Mathematical Statistics and Probability*, 1961. 5.3.3

[150] Brian Rogers and Maureen Graham. Motion parallax as an independent cue for depth perception. *Perception*, 8(2):125–134, 1979. 1.1

[151] Brian Rogers and Maureen Graham. Similarities between motion parallax and stereopsis in human depth perception. *Vision research*, 22(2):261–270, 1982. 1.1

[152] Anirban Roy and Sinisa Todorovic. Monocular depth estimation using neural regression forest. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2016. 5.1

[153] Aswin C Sankaranarayanan, Ashok Veeraraghavan, Oncel Tuzel, and Amit Agrawal. Specular surface reconstruction from sparse reflection correspondences. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1245–1252. IEEE, 2010. 1.3.3

[154] Zoltán Sárosi, Wolfgang Knapp, Andreas Kunz, and Konrad Wegener. Detection of surface defects on sheet metal parts by using one-shot deflectometry in the infrared range. In *Infrared Camera Applications Conference 2010*, pages 243–254. ETH Zurich, IWF, 2010. 1.3.3

[155] Charles Saunders, John Murray-Bruce, and Vivek K Goyal. Computational

periscopy with an ordinary digital camera. *Nature*, 565(7740):472, 2019. 3.1

[156] Ashutosh Saxena, Sung H. Chung, and Andrew Y. Ng. Learning depth from single monocular images. *Advances in Neural Information Processing Systems*, 2006. 5.1

[157] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 2002. 5.1

[158] ANNELI Schulz, ROMAN KROBOT, EVELYN OLESCH, CHRISTIAN FABER, FRIEDRICH STINZING, CHRISTIAN STEGMANN, GERD Hausler, et al. Methods for the characterization of mirror facets for imaging atmospheric cherenkov telescopes. In *Proc. ICRC*, volume 32, pages 1–4, 2011. 1.3.3

[159] James A Sethian. Fast marching methods. *SIAM review*, 41(2):199–235, 1999. 2.2.1

[160] James A Sethian. *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*, volume 3. Cambridge university press, 1999. 2.2.1

[161] James A Sethian and A Mihai Popovici. 3-d traveltime computation using the fast marching method. *Geophysics*, 64(2):516–523, 1999. 3.5

[162] Jianping Shi, Xin Tao, Li Xu, and Jiaya Jia. Break ames room illusion: depth from general single images. *ACM Transactions on Graphics*, 2015. 5.1

[163] Jianping Shi, Li Xu, and Jiaya Jia. Just noticeable defocus blur detection and estimation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2015. 5.1

[164] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgbd images. *European Conference on Computer Vision*, 2012. 5.1

[165] Brandon M. Smith, Matthew O'Toole, and Mohit Gupta. Tracking multiple objects outside the line of sight using speckle imaging. In *CVPR*, 2018. 3.1

[166] Michael Spivak. *Calculus on manifolds: a modern approach to classical theorems of advanced calculus*. CRC Press, 2018. A.5

[167] Pratul P. Srinivasan, Rahul Garg, Neal Wadhwa, Ren Ng, and Jonathan T. Barron. Aperture supervision for monocular depth estimation. *IEEE/CVF Conference on*

Computer Vision and Pattern Recognition, 2018. 5.1

[168] Orestes Stavroudis. *The optics of rays, wavefronts, and caustics*. Elsevier, 1972. 2.3, 2.4.1, 4.3.1, 4.3.3

[169] Supasorn Suwajanakorn, Carlos Hernandez, and Steven M. Seitz. Depth from focus with your mobile phone. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2015. 5, 5.1

[170] Rick Szeliski and Polina Golland. Stereo matching with transparency and matting. *International Journal of Computer Vision*, 1999. 5

[171] Matthew Tancik, Guy Satat, and Ramesh Raskar. Flash photography for data-driven hidden scene recovery. *arXiv preprint arXiv:1810.11710*, 2018. 3.1

[172] Huixuan Tang and Kiriakos N. Kutulakos. Utilizing optical aberrations for extended-depth-of-field panoramas. *Asian Conference on Computer Vision*, 2012. 5.3.3, 5.4.2, 5.4.4

[173] Huixuan Tang, Scott Cohen, Brian Price, Stephen Schiller, and Kiriakos N. Kutulakos. Depth from defocus in the wild. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2017. 1.1, 1.3.4, 5.1

[174] Christos Thrampoulidis, Gal Shulkind, Feihu Xu, William T Freeman, Jeffrey Shapiro, Antonio Torralba, Franco Wong, and Gregory Wornell. Exploiting occlusion in non-line-of-sight active imaging. *IEEE Transactions on Computational Imaging*, 2018. 3.1

[175] Chia-Yin Tsai, Kiriakos N Kutulakos, Srinivasa G Narasimhan, and Aswin C Sankaranarayanan. The geometry of first-returning photons for non-line-of-sight imaging. In *CVPR*, 2017. 3.1, 3.2.1

[176] Chia-Yin Tsai, Aswin C. Sankaranarayanan, and Ioannis Gkioulekas. Beyond volumetric albedo—a surface optimization framework for non-line-of-sight imaging. In *CVPR*, 2019. 3.1, 3.5

[177] Richard Tucker and Noah Snavely. Single-view view synthesis with multiplane images. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 5.3.1, 5.3.1, 5.3.3, D.2

[178] Ashok Veeraraghavan, Ramesh Raskar, Amit Agrawal, Ankit Mohan, and Jack Tumblin. Dappled photography: Mask enhanced cameras for heterodyned light

fields and coded aperture refocusing. *ACM Transactions on Graphics*, 2007. 5.1

[179] Andreas Velten, Thomas Willwacher, Otkrist Gupta, Ashok Veeraraghavan, Moungi G Bawendi, and Ramesh Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. 2012. 3.1, 3.5, 6.1

[180] Andreas Velten, Di Wu, Adrian Jarabo, Belen Masia, Christopher Barsi, Chinmaya Joshi, Everett Lawson, Moungi Bawendi, Diego Gutierrez, and Ramesh Raskar. Femto-photography: Capturing and Visualizing the Propagation of Light. *ACM Trans. Graph.*, 32(4):44:1–44:8, July 2013. ISSN 0730-0301. 3.1

[181] Neal Wadhwa, Rahul Garg, David E. Jacobs, Bryan E. Feldman, Nori Kanazawa, Robert Carroll, Yair Movshovitz-Attias, Jonathan T. Barron, Yael Pritch, and Marc Levoy. Synthetic depth-of-field with a single-camera mobile phone. *ACM Transactions on Graphics*, 2018. 5.1, 5.4.4, 5.9(i), **??**, D.2(i)

[182] Bruce Walter, Shuang Zhao, Nicolas Holzschuch, and Kavita Bala. Single scattering in refractive media with triangle mesh boundaries. In *ACM Transactions on Graphics (TOG)*, volume 28, page 92. ACM, 2009. A.5

[183] Ting-Chun Wang, Manmohan Chandraker, Alexei A Efros, and Ravi Ramamoorthi. Svbrdf-invariant shape and reflectance estimation from light-field cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5451–5459, 2016. 1.2.1

[184] Yilun Wang, Junfeng Yang, Wotao Yin, and Yin Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 2008. 5.1

[185] Masahiro Watanabe and Shree K. Nayar. Rational filters for passive depth from defocus. *International Journal of Computer Vision*, 1998. 1.1, 1.3.4, 5.1

[186] Bennett Wilburn, Neel Joshi, Vaibhav Vaish, Eino-Ville Talvala, Emilio Antunez, Adam Barth, Andrew Adams, Mark Horowitz, and Marc Levoy. High performance imaging using large camera arrays. In *ACM SIGGRAPH 2005 Papers*, pages 765–776. 2005. 1.1

[187] Robert J Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):139–144, 1980. 2.3.1

[188] Di Wu, Gordon Wetzstein, Christopher Barsi, Thomas Willwacher, Qionghai Dai,

and Ramesh Raskar. Ultra-fast lensless computational imaging through 5d frequency analysis of time-resolved light transport. *International journal of computer vision*, 110(2):128–140, 2014. 3.5

[189] Feihu Xu, Gal Shulkind, Christos Thrampoulidis, Jeffrey H Shapiro, Antonio Torralba, Franco NC Wong, and Gregory W Wornell. Revealing hidden scenes by photon-efficient occlusion-based opportunistic active imaging. *Optics express*, 26 (8):9945–9962, 2018. 3.1

[190] Yongjia Xu, Feng Gao, and Xiangqian Jiang. A brief review of the technological advancements of phase measuring deflectometry. *PhotoniX*, 1(1):1–10, 2020. 1.3.3

[191] Jiawei Zhang, Jinshan Pan, Wei-Sheng Lai, Rynson W.H. Lau, and Ming-Hsuan Yang. Learning fully convolutional networks for iterative non-blind deconvolution. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2017. 5.1

[192] Yinda Zhang, Neal Wadhwa, Sergio Orts-Escolano, Christian Häne, Sean Fanello, and Rahul Garg. Du^{2}net: Learning depth estimation from dual-cameras and dual-pixels. *European Conference on Computer Vision*, 2020. 5.1

[193] Changyin Zhou and Shree K. Nayar. What are good apertures for defocus deblurring? *IEEE International Conference on Computational Photography*, 2009. 5.1

[194] Changyin Zhou, Stephen Lin, and Shree K. Nayar. Coded aperture pairs for depth from defocus. *IEEE/CVF International Conference on Computer Vision*, 2009. (document), 1.4, 1.1, 5.1, 5.3, 5.3.2, 5.4.4, 5.8(d), 5.9(e), **??**, D.1, D.1, D.1(d), D.2(e)

[195] Tian Zhou, Kun Chen, Haoyun Wei, and Yan Li. Improved method for rapid shape recovery of large specular surfaces based on phase measuring deflectometry. *Applied optics*, 55(10):2760–2770, 2016. 1.3.3

[196] Tinghui Zhou, Matthew Brown, Noah Snavely, and David G. Lowe. Unsupervised learning of depth and ego-motion from video. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2017. 5.1

[197] Tinghui Zhou, Richard Tucker, John Flynn, Graham Fyffe, and Noah Snavely. Stereo magnification: Learning view synthesis using multiplane images. *ACM Transactions on Graphics*, 2018. 5, 5.3.1

[198] Andrew Zisserman, Peter Giblin, and Andrew Blake. The information available to a moving observer from specularities. *Image and vision computing*, 7(1):38–42,

1989. (document), 1.3.3, 4, 4.1, 4.2

[199] Yuliang Zou, Zelun Luo, and Jia-Bin Huang. DF-net: Unsupervised joint learning of depth and flow using cross-task consistency. *European Conference on Computer Vision*, 2018. 5.1