SHAPE ESTIMATION UNDER GENERAL REFLECTANCE AND
TRANSPARENCY

by

Nigel J. W. Morris
Dept. of Computer Science
University of Toronto

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Computer Science
University of Toronto

Abstract

Shape Estimation under General Reflectance and Transparency

Nigel J. W. Morris

Dept. of Computer Science

University of Toronto

Doctor of Philosophy

Graduate Department of Computer Science

University of Toronto

2011

In recent years there has been significant progress in increasing the scope, accuracy and flexibility of 3D photography methods. However there are still significant open problems where complex optical properties of mirroring or transparent objects cause many assumptions of traditional algorithms to break down.

In this work we present three approaches that attempt to deal with some of these challenges using a few camera views and simple illumination.

First, we consider the problem of reconstructing the 3D position and surface normal of points on a time-varying refractive surface. We show that two viewpoints are sufficient to solve this problem in the general case, even if the refractive index is unknown. We introduce a novel "stereo matching" criterion called *refractive disparity*, appropriate for refractive scenes, and develop an optimization-based algorithm for individually reconstructing the position and normal of each point projecting to a pixel in the input views.

Second, we present a new method for reconstructing the exterior surface of a complex transparent scene with inhomogeneous interior. We capture images from each viewpoint while

moving a proximal light source to a 2D or 3D set of positions giving a 2D (or 3D) dataset per pixel, called the *scatter-trace.* The key is that while light transport within a transparent scene's interior can be exceedingly complex, a pixel's scatter trace has a highly-constrained geometry that reveals the direct surface reflection, and leads to a simple "Scatter-trace stereo" algorithm for computing the exterior surface geometry.

Finally, we develop a reconstruction system for scenes with reflectance properties ranging from diffuse to specular. We capture images of the scene as it is illuminated by a planar, spatially non-uniform light source. Then we show that if the source is translated to a parallel position away from the scene, a particular scene point integrates a magnified region of light from the plane. We observe this magnification at each pixel and show how it relates to the source-relative depth of the surface. Next we show how calibration relating the camera and source planes allows for robustness to specular objects and recovery of 3D surface points.

Dedication

This work is dedicated to my family who have supported me through all the highs and lows.

Contents

Chapter 1

Introduction

"Vision is the art of seeing what is invisible to others."

-Jonathan Swift

One of the core elements of the human visual system is "depth perception," which enables us to analyze our world and perform many tasks which would otherwise be extremely challenging or impossible. Depth perception is also an important component in artificial vision systems, however certain optical properties of everyday scenes often make this task much more difficult. While the human visual system is generally qualitatively robust to optics involving mirrors, transparency and scattering media, these properties often confound artificial systems. While artificial systems have much greater quantitative precision, we would like to be able to add the same kind of robustness as in biological systems.

The 3D scanning of objects for synthesis in virtual environments is an important motivation for this work that requires both robustness and quantitative accuracy. The quality of rendering or simulation of virtual environments has advanced to such a degree that one of the primary limiting factors in virtual images is the detail in the geometry. Many hours are re-
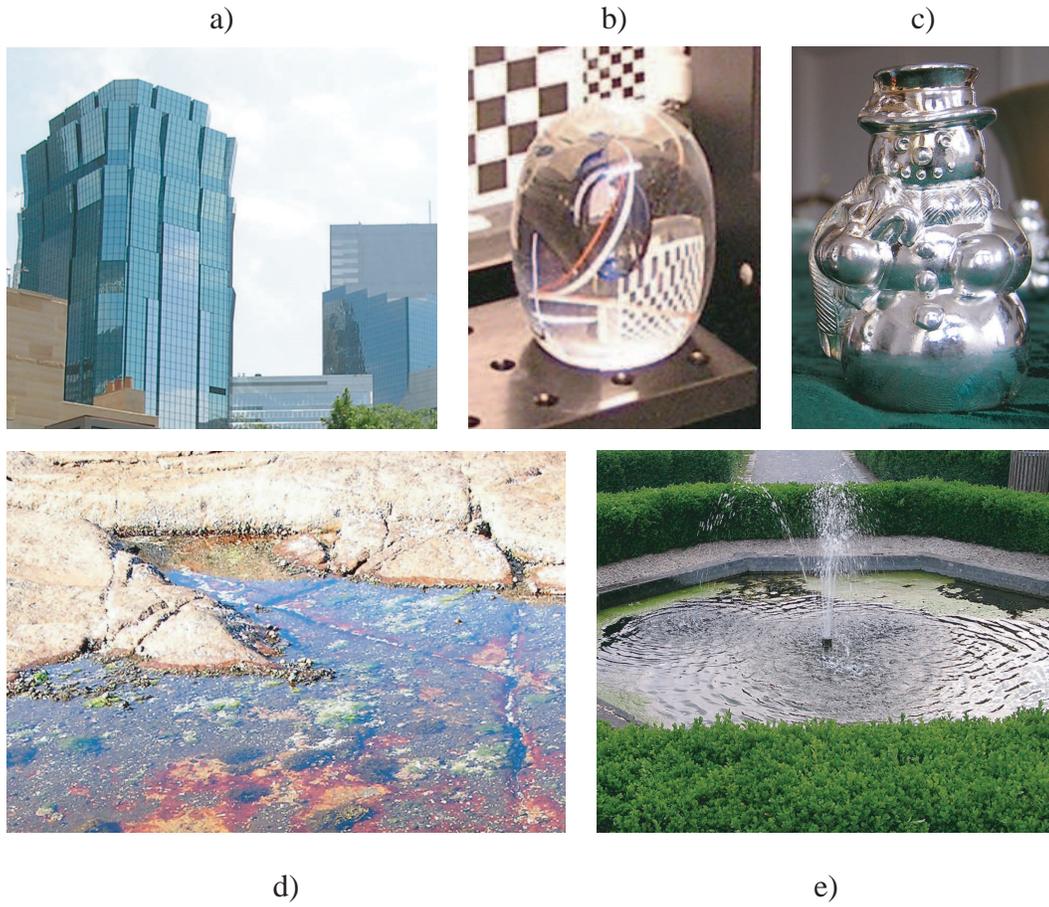
Figure 1.1: Everyday example scenes that have challenging optical properties. Mirror reflection: all. Transparency: b), d) and e)

quired to manually generate the required detail to match realism, so it makes more sense to scan the geometry from the real world. Real world objects often present complex optics and transparency making it important for scanning techniques to be able to deal with such phenomena. This same scanning procedure is also vital for preserving the world's cultural heritage, as fragile archaeological artifacts can be preserved in this manner and virtual museums can allow people across the globe a unique perspective on such artifacts. Another potential application of such vision systems is in quality assurance and inspection for the manufacture of glass or plastic products that exhibit transparency. For instance, glass flow regulators for intravenous machines currently must be painstakingly and manually examined to ensure that no defects are present.

One of the major challenges in trying to scan such objects is that they do not have intrinsic appearance, but rather they redirect the incident light from the background either through reflection of refraction. This makes them very difficult to scan with a traditional laser scanner without resorting to special optics [7], immersing them [52, 123] or directly coating the objects [37]. While simply coating such objects with a paint or powder that has much simpler reflectance properties might appear to be a good work-around, this presents several problems in itself. First, this requires potentially undesirable contact with the target objects, and objects that are precious artifacts or glass may be extremely fragile and the coating process may easily damage them. Second, if the objects do have some colored or textured parts these will be obscured by a general coating. Finally, the coating itself will subtly change the shape of the object and will tend to fill in cracks or nooks, artificially smoothing the shape. Without great care coating can easily be uneven and add spurious bumps or layers.

Despite its chemical simplicity, water exhibits many of the most challenging optical properties and both oceanography [23, 58, 64, 136] and fluid mechanics [41] are important domains

that would benefit from accurate and rapid perception of water and waves. It is also important for robots to be able to navigate scenes that may exhibit such properties. For instance, the recent development of rescue robots for people trapped in hazardous environments could certainly benefit from navigating around broken glass or seeing through flooded environments.

The problem of capturing the shape of scenes has been a primary focus of computer vision research. Initial work predominately made significant simplifying assumptions about the reflectance properties of the scene, generally assuming a simple diffuse reflectance model [105, 109]. More recently, attempts have been made to generalize to scenes with more arbitrary reflectance models [45, 61, 122, 125, 140, 142]. While these advances are important, there has been little work on dealing with even more general scenes that may contain both opaque and non-opaque structures as well as a variety of reflectance properties and indirect light transport.

1.1 Contributions

The goal of this work is to show that 3D photography methods can be used to recover the shape of general scenes containing transparent and shiny objects by making use of controlled diffuse illumination patterns and simple multi-view camera systems.

Here is a summary of specific contributions made in this thesis. It includes expanded versions of some work that has previously been published [84, 85].

- A system for finding stereo correspondences at each frame of a temporally dynamic refractive liquid. These correspondences can then be used to recover the 3D points and normals on the surface as well as its refractive index. We can then display novel views using the recovered geometry.

- A system for reconstructing 3D points and normals of glass and specular surfaces by

modeling direct reflection in order to be invariant to indirect light that may be internally transmitted or caused by inter-reflections.

- A fast method for recovering the depth of scenes with general reflectance properties from a monocular view with simple active illumination.

1.2 Overview

In the next chapter we cover the necessary background material for this work. We discuss the types of geometric representations that are often reconstructed as well as a discussion of *ray optics* and the physical and optical properties of scenes. The next chapter covers our work on the 3D reconstruction of dynamic transparent surfaces. Next we present our approach to recovering the exterior surface of reflective and inhomogeneous transparent surfaces. Then we cover our work that uses a related illumination method to recover depth from general scenes from a single camera. Finally we present our conclusions.

Chapter 2

Background

"Never bring the problem solving stage into the decision making stage. Otherwise, you surrender yourself to the problem rather than the solution."

-Robert H. Schuller

There has been significant progress in 3D photography and image-based measurement techniques in recent years. The shape of more and more complex scenes can be measured to higher precision. In this chapter we present and categorize many of these advances as well as examine what open areas and unanswered questions remain. In this work we are primarily concerned with the 3D reconstruction of surfaces. Given this, we will first examine the various ways that surfaces can be geometrically represented as well as their physical interaction with light.

2.1 Geometric Representations

2.1.1 Depths

The first and most basic representation of shape is in the form of depth maps. For a given image of the scene, we define *pixel rays* as vectors originating at the camera center of projection and passing through the center of each pixel into the scene. We assume a simple pinhole camera projection model. A depth map consists of an assignment to every pixel of depth, that is the distance from the center of projection to the first intersection of the pixel ray with a surface in the scene (see Figure 2.1).

Depth maps are often the output of stereo systems where corresponding features are found in the images taken from two or more viewpoints of the scene [105]. These correspondences define rays from each center of projection into the scene that converge at a particular depth. The depth map is created from the depths assigned to pixels of a particular 'reference' image selected from the available images.

2.1.2 Normals

Normals are defined as vectors perpendicular to the tangent plane at a surface point. A normal map consists of an assignment to each pixel of the normal at the surface point projecting to the center of each pixel (see Figure 2.1).

Normal maps can be obtained by reconstruction methods that take into account the shading of an object such as photometric stereo [131] or shape from shading [51].

A normal map models the surface as locally planar facets however it is often more accurate to measure the local surface projecting to a pixel as a collection of planar micro-facets, each with their own normal [88]. These normals can be collected into a histogram of directions and

can be used to better represent rough materials.

2.1.3   Point Clouds and Surfels

While the previous two representations were image specific, *point clouds* and *surfels* represent the surface of an object as a disconnected "cloud" of 3D points possibly with associated surface normals (in the case of surfels). Point clouds may be obtained from sparse multi-view reconstruction approaches that are not constrained to a particular reference image [35]. They are also the typical output of laser scanners. Reconstructing the shape of surfaces from a point cloud is a challenging and open problem [60]. The additional normal information stored in surfels allows for improved reconstruction algorithms especially when the frequency of noise in the normals data is different from the point data. For instance, 3D points often contain high frequency noise, while normals often exhibit lower frequency noise. By combining the two we can achieve significantly improved results [63, 91].

2.1.4   Meshes

The previous representations defined the local position and orientation of the surface at distinct points, whereas *meshes* are a piecewise linear representation of the surface. Meshes consist of *vertices* which are 3D points that approximately lie on the surface. The vertices are connected by lines called *edges* and edges that are coplanar form polygonal *faces*. Meshes are useful for visualization, surface smoothing [49] and normal estimation when explicit normals are unknown [114].

Depth Map

Normal Map

Point Cloud

Points and Normals

Surfels

Mesh
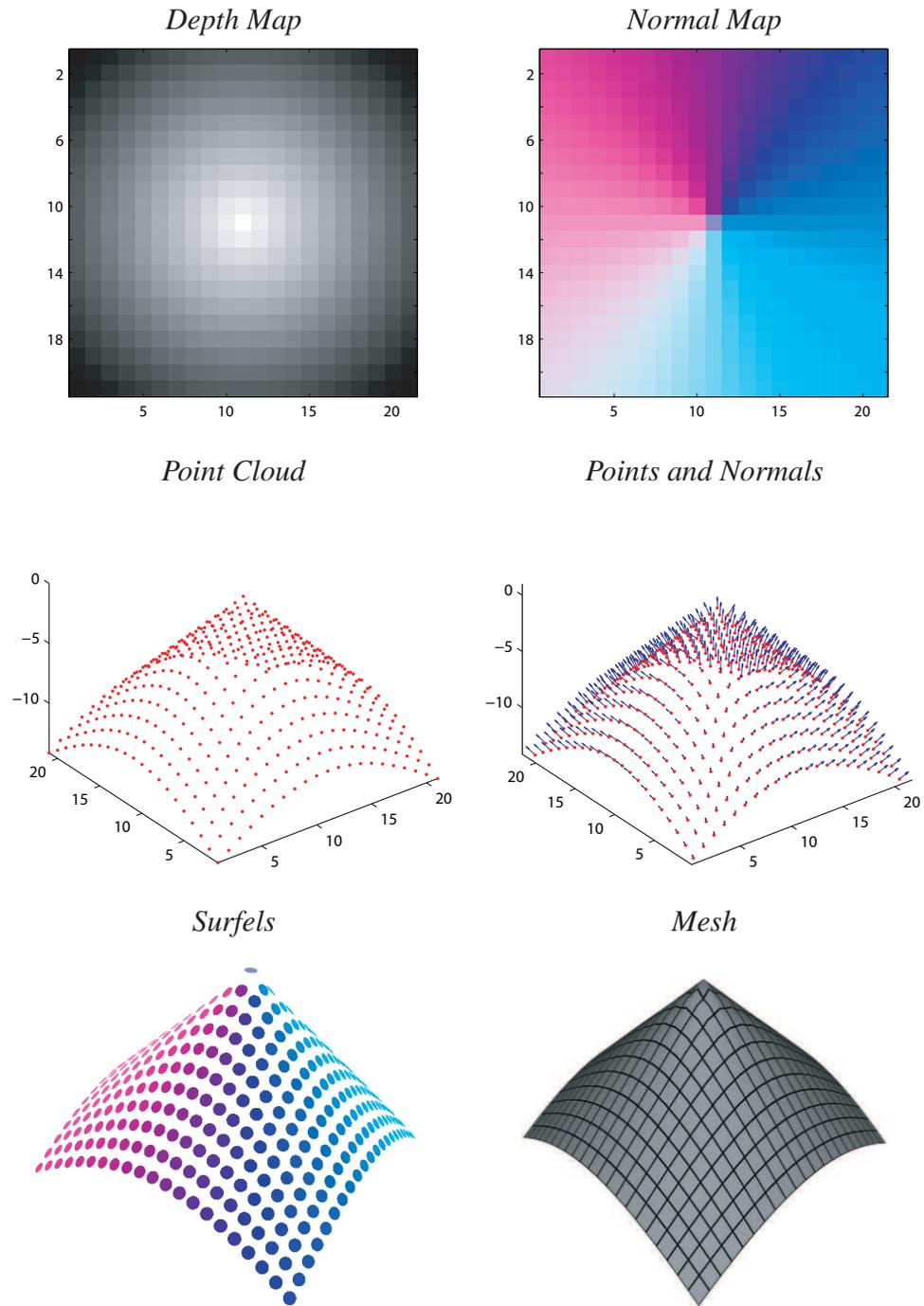
Figure 2.1: Examples of different geometric representations for a cone.

2.2 Physical Properties and Models

2.2.1 Radiometry

Radiometry deals with the measurement of electromagnetic radiation including visible light. We first define the terms and physical quantities involved:

Radiant energy The energy Q_r emitted from or absorbed by a surface during time interval t and measured in Joules (J).

Power The radiant power Φ is the rate of energy emission or absorption and is measured in Watts (W) or Joules per second (Js^{-1}):

$$\Phi = \frac{dQ_r}{dt}.\qquad(2.1)$$

Incident irradiance The power received per unit surface area E, measured in Watts per meter squared (Wm^{-2}).

Solid angle The ratio of a portion S of the surface area of a sphere to the squared radius of the sphere, measured in steradians (sr).

Radiance The rate of emission or absorption at a surface per unit foreshortened area in particular direction. Radiance L is measured in Watts per steradian per meter squared ($Wm^{-2}sr^{-1}$). For example, to measure radiance at an infinitesimal patch at $\mathbf{p}$ on a surface, with incident light coming from an emitting surface patch, we would first project this patch onto the point's hemisphere to find the solid angle. Then we need a cosine term to govern the angular foreshortening. So the radiance $L(\mathbf{p}, \omega)$, at point $\mathbf{p}$ and coming in direction ω, is then given by:
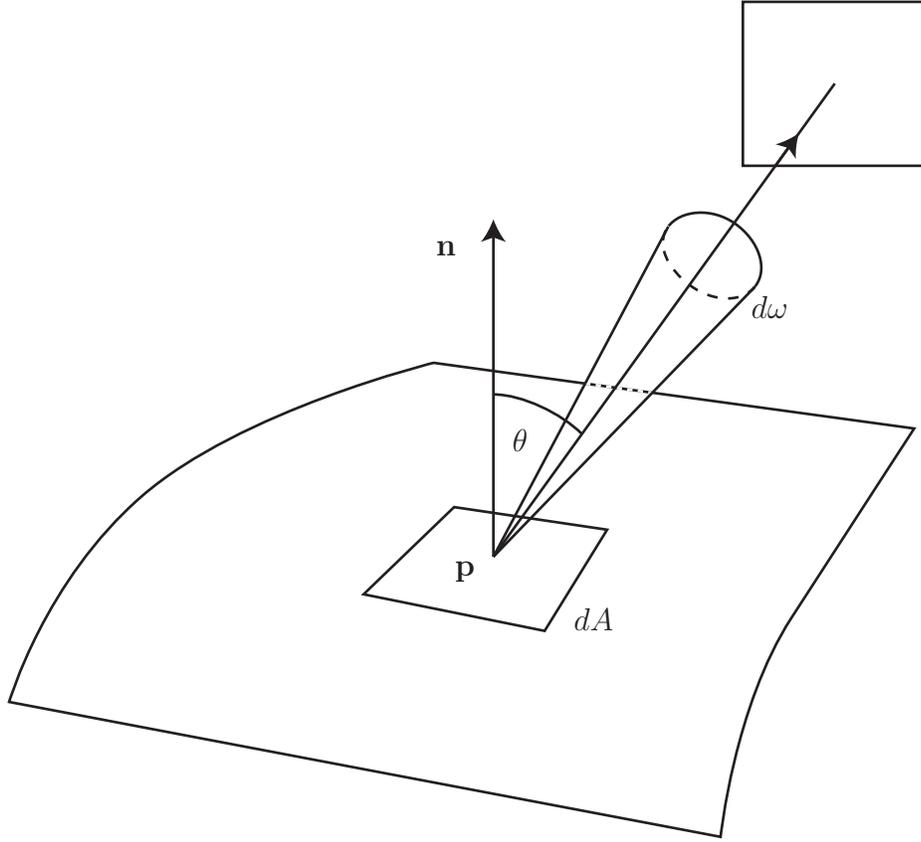
Figure 2.2: Geometrical properties of radiance.

$$L(\mathbf{p}, \omega) = \frac{d^2\Phi}{\cos\theta d\omega dA},$$ (2.2)

where $d\omega$ is the solid angle of the projected patch, θ is the angle between the surface normal and the incident direction, and dA is the area of the patch around $\mathbf{p}$. See Figure 2.2 for the geometry. Irradiance can also be expressed in terms of an integration of radiance over the hemisphere of directions:

$$E(\mathbf{p}) = \int_{\Omega^+} L(\mathbf{p}, \omega) \cos\theta d\omega.$$ (2.3)

2.2.2   Bidirectional Reflectance Distribution Functions

Reflection of light on most opaque surfaces can be described by the bidirectional reflectance distribution function (BRDF) [92]. This function relates incident irradiance to exitant radiance for any pair of directions as follows:

$$f_r(\mathbf{p}, \hat{\omega}_i, \hat{\omega}_o) = \frac{dL(\mathbf{p}, \hat{\omega}_o)}{L(\mathbf{p}, \hat{\omega}_i) \cos \theta_i d\omega_i}. \tag{2.4}$$

The BRDF is the quotient of the exitant radiance $L(\mathbf{p}, \hat{\omega}_o)$ from direction $\hat{\omega}_o$ over the incident irradiance from direction $\hat{\omega}_i$ at the patch about $\mathbf{p}$ measured in sr^{-1}. The BRDF in general is a function of four angular dimensions, two each for the incident and exitant directions. Note that the two directions are constrained to be within the hemisphere around the surface normal.

We can reduce the angular dimensionality to three in the case of isotropy, where there is rotational symmetry in the BRDF around the surface normal. BRDFs of this type are called *isotropic* as opposed to *anisotropic*.

Another property of BRDFs is Helmholtz reciprocity [128], it states that the roles of incident and exitant directions can be reversed because of the second law of thermodynamics. In particular, the following equation holds for any BRDF:

$$f_r(\mathbf{p}, \hat{\omega}_i, \hat{\omega}_o) = f_r(\mathbf{p}, \hat{\omega}_o, \hat{\omega}_i). \tag{2.5}$$

BRDFs are also subject to the law of energy conservation: the reflected energy must not exceed that which is received.
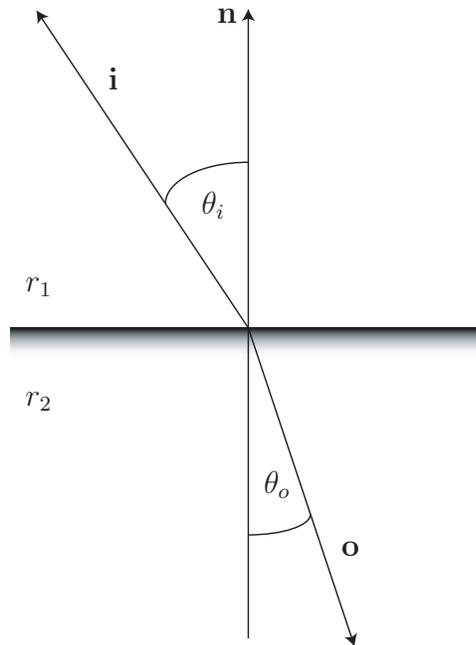
Figure 2.3: Refraction at an interface.

2.2.3 Snell's Law

So far we have only considered opaque surfaces, however transparent materials allow the transmission of light through a surface. In this case, the direction of travel depends on the ratio of the refractive indices on either side of the surface [10]. This refraction occurs in the plane defined by the direction of incidence and the surface normal and the angle of refraction is described by Snell's law:

$$r_1 \sin \theta_i = r_2 \sin \theta_o, \tag{2.6}$$

where r_1 and r_2 are the refractive indices of the two media, θ_i is the angle between the incident direction and the surface normal, and θ_o is the angle between the transmitted direction and the surface normal (Figure 2.3).

Refractive indices depend on wavelength. As a result, refraction causes spectral separation or dispersion of chromatic light. This is often demonstrated by the transmitted 'rainbow' of colors as a beam of white light is refracted through a prism.

2.2.4   Reflection Models

Reflection at a surface depends greatly on the material and fine scale surface structure. We can categorize most types of reflection into three broad types (see Figure 2.4):

Diffuse reflection This model describes a scattering process that causes light to be reflected evenly in all directions. This type of reflection is also known as Lambertian reflectance [71]. Materials that are rough at the micro scale tend to fit this type of model however significant roughness can be more accurately modeled by the Oren-Nayar reflectance model [95].

Specular reflection When incident light strikes a polished surface and it reflects across the surface normal such that the outgoing angle is equal to the incident angle, we call this specular reflection. Mirrors and most transparent objects exhibit strong specular reflection.

Glossy reflection Many materials exhibit a specular lobe of reflection around the specular direction, rather than an impulse. This is caused by fine scale surface irregularities on an otherwise smooth surface and we call such materials 'glossy.'

There are many more specific types of reflection such as retroreflection and subsurface scattering [72, 81]. Many computer vision algorithms are specific to a particular type of surface
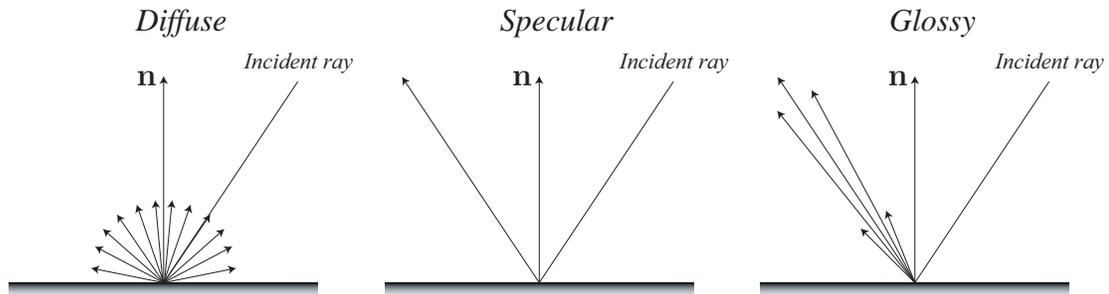
Figure 2.4: Illustrations of the different reflectance models.

reflectance model, however some materials have BRDFs that do not easily fit into these cate-gories and can cause 3D reconstruction algorithms problems since they violate the algorithm's underlying assumptions about the scene's BRDFs.

2.3   Appearance Modeling

In this section we examine techniques that extract radiometric or material properties from im-ages without reconstructing the shape of the target scene.

2.3.1   Environment Matting and Extensions

Environment matting techniques were designed to capture the appearance of specular and trans-parent objects by blending a foreground opacity matte with the background in those areas that are transparent or reflective. These methods model the reflected or transmitted appearance without computing the foreground object's shape. In general, a sequence of patterns is dis-played around the object so that the background regions that map to each pixel by either re-fraction of reflection can be determined [143]. This process was extended to allow for dynamic

transparent objects [21] as well as to remove the necessity of the calibrated camera and patterns [130]. Another extension allows for the simultaneous capture of environment mattes, the approximate shape and reflectance properties of transparent objects [82].

2.3.2   Separating Reflection Components

Diffuse reflection can be separated from specular reflection under certain conditions such as color differences between specular highlights and the diffuse scene [80, 118], polarization [126] and stereo constraints [75]. By identifying the diffuse component we can use simpler algorithms to process a scene as if only diffuse materials were present.

The direct reflection component has also been iteratively separated from light that has undergone multiple reflections with the scene. This can be an important preprocessing step for 3D reconstruction since most algorithms assume only direct reflection [90]. More recently direct reflection has been separated from more general indirect light that also includes sub-surface scattering and other volumetric effects [89].

The reflection component has also been separated from the transmitted component for images of planar transparent objects [73, 106]. This decomposition is strongly under-constrained so significant assumptions such as images with different polarization [106], natural image statistics [73] or user interaction are required.

2.4   Shape Reconstruction of Diffuse Scenes

While the methods described in the previous sections capture the specific radiometric properties of the scene or decompose images into reflectance components, they do not directly measure the scene shape. In this section we review algorithms for recovering scene shape, beginning

with algorithms that work under simple reflectance models and continuing with those that deal with more complex ones.

In order to recover scene shape, we need a model of how light reflects off of the surface of objects. Under the diffuse reflection model the outgoing radiance of a particular surface patch does not vary with viewpoint. What is also often implicitly assumed is that objects in the scene are opaque and indirect lighting such as inter-reflections or subsurface scattering are treated as insignificant in order to simplify the reconstruction algorithms.

Methods such as stereo reconstruction often rely on this model to find correspondences in two or more images of a scene point. Given such a correspondence, triangulation allows for depth reconstruction. A review of the developments in two view stereo is available in [105]. Multi-view methods generalize this approach to include data from more views and have been extensively researched [109].

While these methods do an admirable job of reconstructing shape, everyday scenes frequently violate the diffuse reflectance model. In the following sections we will look at work that applies to broader classes of materials.

2.5   Shape Reconstruction of Non-Diffuse Scenes

Simple diffuse reflection models are insufficient for 3D reconstruction of general scenes with complex reflectance properties. Rather than performing a layer decomposition or treating glossy or specular highlights as outliers [11, 134], it is possible to take advantage of optical and geometric constraints to perform 3D reconstructions on such scenes.

2.5.1   Multi-view Approaches

The reciprocity of BRDFs (Section 2.2.2) has been utilized to enable stereo on scenes with general BRDFs [79, 125, 141, 142]. These approaches use Helmholtz stereopsis, where the positions of a point light source and camera are interchanged to obtain stereo pairs such that the appearance of corresponding image points are invariant to the BRDF of scene points. This of course does not deal with indirect light, and specular points would still be challenging to recover since they would not reflect light significantly in non-specular directions.

Another approach takes advantage of the fact that for light incident at a patch from a fixed direction and for a fixed outgoing direction, the reflected radiance scales proportionally to the power of the light source. So under these conditions, if the source power is multiplied by a factor k, the outgoing radiance also scales by the same factor. In [24], a stereo constraint is presented where a scene under non-isotropic illumination is observed by two cameras and the illumination power is varied by a factor k. This can be accomplished with a non-isotropic, directional source that is rotated between images and this leads to a stereo constraint that is used to reconstruct scene depth. The advantage of this approach is that it is independent of the BRDF, however the method fails for highly glossy and mirror-like specular materials where the constraint can only be applied to points that specularly reflect light toward the camera.

2.5.2   Photometric Stereo and Monocular Approaches

Example-based photometric stereo methods rely on an orientation-consistency invariant to reconstruct objects with arbitrary BRDFs. Orientation consistency states that, under illumination from a distant point source and orthographic viewing, two surface points with the same surface normal and the same BRDF result in identical image irradiance [45, 46, 122]. To reconstruct

an object of unknown shape we add to the scene a reference object with known shape and the same BRDF as the target object, and capture images from many light source positions. We accomplish this by finding correspondences between the example object and the target object where the irradiance is the same in all input images. Thus, by the orientation-consistency invariant, the normal on the unknown object can be copied from the normal of the example object producing a normal map for the unknown object. This work was extended to deal with self-shadowing in [122] and to allow multiple BRDFs on the target object [46]. The approach is limited by the strong orthographic viewing assumption; again, specular objects are challenging to reconstruct for the reasons mentioned in Section 2.5.1.

In [48], BRDF slices are acquired from a fixed viewpoint by varying the illumination source position. Then surface normals and tangent vectors are found photometrically, by taking advantage of symmetry in the BRDF slices. While this method is not completely BRDF-invariant, a wide variety of materials can be dealt with. In [5] photometric stereo is extended to estimate the surface normals of spatially varying isotropic materials. The technique utilizes the inherent symmetry of isotropic BRDFs, and does not apply to more general, potentially anisotropic BRDFs that may be spatially varying. Photometric stereo has also been extended to deal with glossy surfaces by taking advantage of light ray direction cues from the target object's shadow [22].

Based on the observation that color space transformations can be used to effectively separate diffuse and specular reflection [80], various color space invariants have been used to reconstruct some non-diffuse scenes [140]. The performance of these approaches depends on the angular difference in the color space between the diffuse and specular colors.

2.6 Shape Reconstruction of Specular Scenes

While many of the techniques described in the previous section are not suitable for reconstructing perfectly specular objects, there is a significant body of work that has been dedicated to the reconstruction of such scenes. Perceptual analysis of specular scenes shows that the movement of specular highlights improves human observers' shape discrimination [32, 57]. This suggests that there are shape cues available to artificial systems beyond the standard shading cues [51].

2.6.1 Highlight Tracing

A highlight is a distorted reflection of a point light source on a specular surface. Stereoscopic observations of specular highlights have been used to infer surface curvature information [12]. Highlights have also been used to iteratively grow the shape of specular surfaces from known "seed points" on the surface using several images from different viewpoints [108]. Sanderson showed how specular highlights could be used to recover sparse depths on the surface using stereo views, showing that with a single view there is an inherent ambiguity between surface orientation and depth. More recently, highlights of a known light source on a specular surface were tracked and used to iteratively update a hypothesized model of the specular object's surface [113]. One of the main drawbacks of these methods is that they deal with point light sources. This makes it is difficult to get broad coverage of the surface since highlights are generally very localized. In [137], a toroidal light source was used to create extended highlights on specular objects and then used to recover depth as the object was rotated relative to the source and camera. While this improves coverage, many images are still required as the object is rotated.

Recently, dense specular flow caused by relative motion between a smooth specular object

and its surrounding, unknown environment has been shown to enable shape estimation of the surface by solving a PDE [1]. This has since been simplified by combining several distinct flows that allow for a linear solution [127].

The analysis of specular highlights has also been applied to glossy surfaces in [18, 33]. A dense sampling of highlights was used in [18] and in [33] a polarizing filter was used to isolate the specular component and recover surface orientation. A related approach for recovering fine scale depths of specular or partially specular objects, [129], uses a parabolic mirror to focus coaxially aligned illumination rays and viewing rays to a point on the target surface. The incident illumination is varied over the surface of the mirror, covering a range of incidence angles and the response measured. A dense normal map is found over the surface by computing the bisectors of the incident and outgoing rays for points on the surface as the position of the mirror is shifted to scan the surface.

2.6.2   Laser Scanning of Specular Scenes

Laser range-finders have also been adapted to reconstruct the shape of specular objects [7]. This is done by restricting the angle of the incident light to a single direction by attaching several parallel plates at an angle in front of the CCD elements. The vertical plates, along with a horizontal slit collimate the incident light to a narrow range of angles, allowing approximate surface triangulation. The problem with this approach is that the object must be rescanned many times from different angles to capture all of the surface normals present on the object.

2.6.3 Dense Calibrated Pattern Approaches

More recent work has approached the problem by using dense, calibrated patterns placed so that their reflection can be observed from the camera. Knowing the 3D location of points on the reflected pattern is still insufficient to overcome the depth/normal ambiguity from a single view (Section 2.6.1), since there are a family of depths and normals that could explain the given reflection. This problem is solved by either moving the pattern to a second position, constraining the reflected rays, or by using a second view to verify hypothesized depths [15, 16, 62, 69, 119]. These techniques are advantageous because they recover dense depth and normal maps and make no assumptions about surface continuity.

By calibrating a dense geometric pattern of lines, it is possible to recover the shape as well as curvature properties of a specular object that reflects the pattern by observing the distortion this reflection causes [103].

In [77] an object was illuminated by a sequence of dense gradient patterns and observed with and without a polarization filter. The addition of the polarization filter allows for both specular and diffuse image normal maps to be created.

2.7 Shape Reconstruction of Transparent Media

While transparent objects are specular in nature and it may appear that many of the specular reconstruction algorithms in Section 2.6 can be directly applied to them, the algorithms were not intended to be robust to the additional optical phenomena associated with transparent objects.

Early work on the reconstruction of refractive surfaces began in photogrammetry [31, 47, 78, 93]. These techniques assume a low-parameter model for a transparent surface (e.g., a plane) and solve a generalized structure from motion problem in which camera parameters,

surface parameters, and 3D coordinates of feature points below the surface are estimated simultaneously.

2.7.1 Tomographic Reconstruction

Transparent objects have been reconstructed using tomography-based approaches that measure the attenuation of light passing through the object [14, 123]. In [123], transparent objects were immersed in a cylindrical column of liquid that was matched to the object's refractive index. Assuming that the object has homogeneous density and is colored, images of the object with back-lighting allow for a tomographic reconstruction. When the object is not colored, a dye can be used in the immersing liquid instead. Again, the object must be homogeneous and any opaque parts would prevent reconstruction. Requiring liquid immersion limits this technique's applicability.

2.7.2 Fluorescent Immersion

In related work [52], transparent objects were immersed in a liquid matching the object's refractive index. The liquid was infused with a dye that fluoresces when illuminated by a laser. Then the object was scanned by sheets of laser light and viewed from a perpendicular direction. Since only the liquid contains the dye and lights up, the transparent objects leave voids in the images and a volume representing the object can be reconstructed. Even if the liquid is not matched in refractive index, the object's surface nearest to the laser scanner can still be recovered.

2.7.3   Polarization-based Reconstruction

When light reflects off of a surface, a fraction of the light becomes polarized in the direction of the surface normal. The phase of the reflected polarized light encodes the orientation of the reflection plane which is defined as the plane spanned by the surface normal and the incident ray [100].

Several methods exist for determining the surface normal once the reflection plane is determined. One technique is to use a second view to constrain the normal to an epipolar line and then use a global minimization approach to solve for the surface normals as well as depths [100].

Another approach assumes surface smoothness and that the normals on the object's silhouette are perpendicular to the viewing angle. Once the silhouette normals are determined, degree of polarization images are used to propagate the reconstruction to the rest of the object [83].

2.7.4   Shape from Distortion

The apparent distortion of the background by a refractive object has led to a number of methods for recovering the shape of the transparent object. One such method for inferring the shape and pose of transparent objects uses a sequence of images of the object from a moving camera [9]. Features are tracked throughout the sequence as they are distorted by the transparent object and an objective function that characterizes the shape and pose of the transparent object is minimized. In order for this procedure to be tractable, the target objects are constrained to single parameter, homogeneous shapes such as super-quadrics.

An alternative approach models the distortion caused by a refracting surface as a multi-perspective projection of the background. In the case of a scene immersed in a planar refractive

known points

transparent
object

light
paths

camera
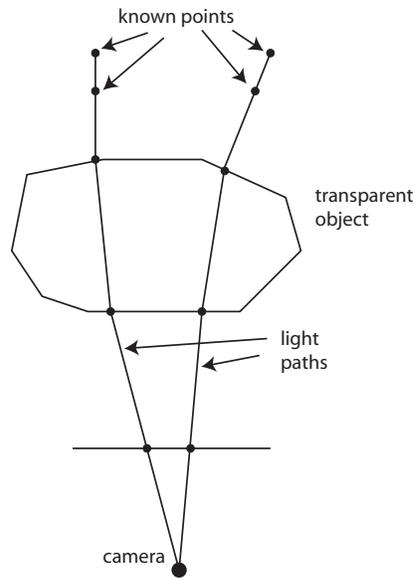
Figure 2.5: Light path triangulation for a transparent object

medium, the projection associated with this as well as a calibration procedure is outlined in
[121].

A framework for reconstructing homogeneous transparent objects was also proposed in [70]
where light rays are traced through the target object and intersected with a pair of known pat-
terns (Figure 2.5). The rays are assumed to have exactly two intersections with the transparent
object, entering and exiting it. Three viewpoints are then needed to reconstruct the 3D positions
and normals of the intersections with the refractive object. While, this work contributes to the
reconstruction of more general transparent objects, it is still limited to homogeneous objects
with only two surface interfaces per ray.

2.7.5 Reconstruction of Dynamic Transparent Media

Shape from distortion techniques have also been applied to the problem of 3D reconstruction of dynamic refractive surfaces, most often for water, whose refractive index is often assumed to be known [3, 59, 87, 93, 132]. In these approaches, the 3D shape of the water surface is recovered by analyzing the distortion of a planar pattern that is immersed in the water. Since a single image is insufficient to fully reconstruct the 3D shape, a variety of assumptions have been used to constrain the problem. These include statistical assumptions about the pattern's appearance over time [87], known average water height [59, 93], and special optics [64, 136]. These assumptions break down when the refractive index is unknown or when the liquid undergoes significant deformations that cause changes in shape and height.

Another approach uses the idea of multi-perspective cameras to model the distortion caused by a refractive surface. This is done by tessellating the target surface into a triangulated mesh and treating each triangle as a linear warping function. The warping can simulate either refraction or reflection and can be modeled as multi-perspective projection [27, 28]. Once the parameters of the warping functions are estimated this leads to measures for the Gaussian and mean curvatures of the surface. While these measurements give important information about the surface, they are insufficient to reconstruct an unambiguous 3D surface.

Yet another technique for reconstructing dynamic transparent surfaces is known as "shape-from-refractive-irradiance" [23, 58, 64, 136]. This approach utilizes a gradient pattern beneath the surface. Light is emitted from this pattern and passes through a collimating lens so that certain intensities or colors correspond to parallel light rays and are then refracted by the surface to the distant camera. This has the result of associating color or intensity with particular surface slopes.

While the tomographic approaches in Section 2.7.1 were constrained to static objects, to-

mographic approaches have also been adapted to work with dynamic transparent liquids [38, 54]. If the liquid is infused with fluorescent dye and viewed through a filter, it can be considered self emitting. Then image irradiance can be related to the distance the observed rays travel within the liquid. This distance is also dependent on the incident angle, since refraction occurs at the liquid surface. Images from multiple viewpoints are used to minimize the error between the expected emission from a hypothesized surface and the actual observations. In [6], dynamic sequences of transparent gas flows with spatially varying refractive index were captured using Schlieren tomography with input from an array of cameras.

2.8   Summary

While the work presented in this chapter represents significant progress toward reconstructing scenes with truly general reflectance properties and transparency, there are still significant gaps remaining. Most of the methods described cannot span the full range of materials from diffuse to fully specular and isotropic to anisotropic. In addition, inter-reflections between objects violate the assumptions of most of the algorithms described above. In some cases this could be overcome by performing a layer decomposition of the direct and indirect light as described in Section 2.3.2, before running the reconstruction algorithms on the direct component. However, this would still be subject to the same shortcomings of the layer decomposition algorithm such as in the case of specularities or caustics.

The current state of the art for 3D reconstruction of transparent scenes has either a limited scope of what can be measured (normal maps, curvature parameters or parametric properties), or special equipment required for scans such as liquid immersion. In addition there are constraints on the properties of target objects, such as homogeneous materials, planar faces, and

number of ray intersections. This raises several questions: How we reduce the constraints imposed by current methods, especially those restricting transparency to homogeneous materials and simple geometric shapes? Can we apply stereo or multi-view techniques to this problem and that of dynamic transparent scenes?

Further measurement of optical properties of transparent objects such as dispersion and refractive index estimation has yet to be fully explored. Finally there has been little work on creating unified reconstruction algorithms for both diffuse and specular materials as well as transparent objects.

Chapter 3

Dynamic Refraction Stereo

> *"Only a fool tests the depth of the water with both feet."*
>
> -African Proverb

3.1 Introduction

Modeling the time-varying surface of a liquid has attracted the attention of many research fields, from computer graphics [21, 30, 76, 86] and fluid mechanics [41] to oceanography [23, 58, 64, 136]. While great strides have been achieved in the development of computer simulators that are physically accurate and visually correct [30, 76], capturing the time-varying behavior of a real liquid remains a challenging problem.

From the point of view of computer vision, analyzing the behavior of liquids from videos poses several difficulties compared to traditional 3D photography applications:

- *No prior scene model:* Spatio-temporal evolution is constrained only by the laws of fluid mechanics, making it difficult to assume a low-degree-of-freedom parametric model for
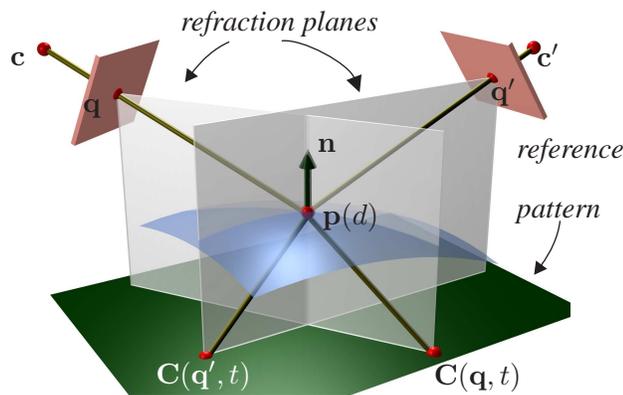
Figure 3.1: Geometry of refraction stereo. The goal is to reconstruct for each pixel $\mathbf{q}$, the 3D position and surface normal of point $\mathbf{p}$ on the refractive surface.

such a scene [9, 135].

- *Non-linear light path:* Liquid surfaces bend the incident light and, hence, a point below the surface will project along a non-linear path to a viewpoint above it.

- *Shape-dependent appearance modulation:* Absorption, scattering and Fresnel transmission cause the appearance of points below the surface to depend on the light's path and, hence, on the surface shape [34].

- *Turbulent behavior:* Liquid flow is an inherently volumetric phenomenon whose complete characterization requires capturing both its time-varying surface and a vector field describing internal motion [138].

- *Instantaneous 3D capture:* Since liquids are dynamic and can flow rapidly, shape recovery must rely on instantaneously-captured information.

As a first step, in this paper we consider the problem of reconstructing the time-varying 3D surface of an unknown liquid by exploiting its refractive properties. To do this, we place a

known, textured pattern below the liquid's surface and capture image sequences of the pattern from two known viewpoints above the liquid (Figure 3.1). Our focus is on imposing as few restrictions as possible on the scene—we assume that the liquid has a constant but unknown index of refraction and that its instantaneous 3D shape is arbitrary, as long as light coming from the pattern is refracted at most once before reaching the input viewpoints.

The reconstruction of refractive surfaces from photographs has a long history in photogrammetry [31, 47, 78, 93]. These techniques assume a low-parameter model for the surface (e.g., a plane) and solve a generalized structure from motion problem in which camera parameters, surface parameters, and 3D coordinates of feature points below the surface are estimated simultaneously. In related work, Treibitz et al. [121], show how a vision system observing a scene immersed under a planar refractive surface becomes multi-perspective. They present a calibration technique to estimate the geometry involved and then recover 3D position of objects immersed in the refractive medium. In computer vision, the reconstruction of time-varying refractive surfaces was first studied by Murase [87], whose seminal work focused on water (whose refractive index is known) and followed a "shape-from-distortion" approach [3, 59, 87, 93, 132]. In this approach, 3D shape is recovered by analyzing one distorted image of a known pattern that is placed underwater. Unfortunately it is impossible, in general, to reconstruct the 3D shape of a general refractive surface from one image, even if its refractive index is known. The inherently ill-posed nature of the problem has prompted a variety of assumptions, including statistical assumptions about the pattern's appearance over time [87], known average water height [59, 93], known points on the surface or surface integrability [120], and special optics [64, 136]. These assumptions break down when the refractive index is unknown or when the liquid undergoes significant deformations that cause changes in shape and height (e.g., pouring water in an empty tank). A different way to approach refractive distortion is to break up the observed sur-

face into a triangulated mesh, where each triangle acts as a general linear camera that warps the background [27, 28]. Solving for the parameters of these GLCs give the Gaussian and mean curvature but the 3D shape remains ambiguous. Another common technique for reconstructing water surfaces is known as "shape-from-refractive-irradiance" [23, 58, 64, 136]. This approach utilizes a gradient pattern beneath the water surface. Light is emitted from this pattern and passes through a collimating lens so that certain intensities or colors correspond to parallel light ray columns and are then refracted by the water surface to the distant camera. This has the result of associating color or intensity with particular surface slopes. This method is most appropriate for measuring small capillary or wind driven waves and is unsuitable for more general use that we are aiming for. Specifically, the lens prevents reconstruction of shallow water as well restricting natural flow in the liquid. Also, the complex nature of an external lens is bound to introduce additional errors due to distortion as well as light attenuation in the liquid. More recent work has focused on tomography based approaches. First, Trifonov et al. [123] immerse target transparent objects in transparent liquid with a matching refractive index and measure the attenuation of a back light through the objects from various views. Then, in [52], objects are scanned by again immersing them in liquid infused with a Fluorescent dye that responds to laser light. When a laser sheet is passed over the object only the dye responds and contours of the object can be recovered.

A closely related problem is the reconstruction of highly specular surfaces such as mirrors [12, 40, 55, 96, 104, 108, 112, 120]. Mirrors interact with light in much the same way that refractive surfaces do—light incident at a point is reflected according to the point's surface normal, thereby tracing a non-linear path. Blake [12] proposed using a moving observer to recover the differential properties of a smooth mirror surface from the observed motion of specularities. Sanderson *et al* [102] were the first to analyze the ambiguities in single-view

mirror reconstruction and to propose a stereo camera configuration for resolving them. Our work, which is based on a novel analysis of two-view ambiguities for refractive scenes, exploits some of the same basic insights. Recently, there have been several approaches that use multiple views to recover the 3D shape of mirror surfaces [15, 16, 62]. While Wang and Dana use a single view, a parabolic mirror allows surface points to be viewed from multiple angles enabling recovery of high resolution relief and texture on specular surfaces [129]. These ideas have also been extended to deal with near-specular and glossy surfaces [18, 33]. In addition, the specular flow of the environment viewed on a mirror surface has been shown to give shape cues even for unknown environments [1, 127].

Reconstructing transparent liquid surfaces is even more challenging than mirrors for three reasons. First, the interaction between light and a mirror does not depend on the mirror's material properties but it does depend on a liquid's refractive index. When this index is unknown, it must be estimated along with 3D shape. Second, the non-linearity of light paths cannot be taken for granted in the case of fluctuating liquid surfaces, whose distance from a pattern below the surface may approach zero, diminishing the effect of refraction. To guarantee stable shape solutions, a reconstruction algorithm must be immune to such degeneracies. Third, establishing accurate pixel-wise correspondences between patterns and their distorted images is much easier in the case of a mirror. In liquids, the distortions are both geometric and radiometric (due to absorption, Fresnel effect, etc.) and can vary significantly from one instant to the next.

The starting point for our work is a novel geometrical result showing that two viewpoints are sufficient to compute both the shape and the refractive index of an unknown, generic refractive surface. The only requirements are (1) knowledge of a function that maps each point on the image plane to a known 3D point that refracts to it, and (2) light is refracted only once. Compared to mirrors, this is a stronger two-view result because it shows that the refractive

index ambiguity, not present in mirror scenes, can be resolved without additional views.

On the practical side, our interest is in algorithms that can capture the detailed dynamic behavior of free-flowing liquids. To this end, our work has four contributions. First, we formulate a novel optimization criterion, called *refractive disparity*, appropriate for refractive scenes, that is designed to remain stable when refraction diminishes. Second, we develop an optimization-based algorithm for individually reconstructing the position and normal of each point projecting to the input views. The algorithm is closer to traditional triangulation [44] and bundle adjustment [43, 124] than to voxel-based stereo [15], and imposes no constraints on the liquid's shape or its evolution. Third, we show that refraction stereo can produce a detailed, full-resolution depth map and a separate, full-resolution normal map for the unknown surface. To our knowledge, only one other shape recovery method, Helmholtz stereopsis [141], has demonstrated the ability to compute dense normal maps along with some depth information (although its depth maps were low resolution and deemed inaccurate). Fourth, we present experimental results for a variety of complex, deforming liquid surfaces. These results suggest that refraction stereo can yield detailed reconstructions that capture the complexity and dynamic behavior of liquids.

3.2   Refraction Stereo Geometry

Consider an unknown, smooth, transparent surface that is viewed by two calibrated cameras under perspective projection (Figure 3.1). We assume that the surface bounds a homogeneous transparent medium (e.g., water or alcohol) with an unknown refractive index. Our goal is to compute the refractive index of the medium and the 3D coordinates and surface normal at each point on the unknown surface. To do this, we place a known reference pattern below the

surface and compute a pixel-to-pattern correspondence function, $\mathbf{C}(\mathbf{q}, t)$, that gives us the 3D coordinates of the point on the pattern that refracts to pixel $\mathbf{q}$ at time t. In the following, we assume that this function is known and concentrate on the instantaneous reconstruction problem at time t. We consider the problem of estimating the correspondence function in Section 3.5. To simplify notation, we omit the time parameter in the following discussion.

Let $\mathbf{q}$ be a pixel in the input views, let $\mathbf{C}(\mathbf{q})$ be the point refracting to $\mathbf{q}$, and suppose that this refraction occurs at distance d from the image plane, at a point $\mathbf{p}(d)$ on the ray through pixel $\mathbf{q}$ (Figure 3.1). The relation between pixel $\mathbf{q}$ and points $\mathbf{C}(\mathbf{q})$ and $\mathbf{p}(d)$ is governed by Snell's law which describes how light is redirected at the boundary between two different media [34]. Snell's law can be expressed as two independent geometric constraints:

- a *deflection constraint*, establishing a sinusoidal relation between incoming and outgoing light directions:

$$\sin \theta_o = r \sin \theta_i \tag{3.1}$$

  where θ_i is the angle between the surface normal and the ray through $\mathbf{C}(\mathbf{q})$ and $\mathbf{p}(d)$; θ_o is the angle between the surface normal and the ray through pixel $\mathbf{q}$; and r is the refractive index;

- and a *planarity constraint*, forcing the surface normal at $\mathbf{p}(d)$ to lie on the plane defined by point $\mathbf{C}(\mathbf{q})$ and the ray through $\mathbf{q}$; we call this plane the *refraction plane* of pixel $\mathbf{q}$.

These two constraints give us a relation between the pixel, a known 3D point that refracts to it, and the unknown surface. Unfortunately, they are not sufficient to determine how far from the image plane the refraction occurs, even when we do know the refractive index. This is because for every hypothetical distance there is a 1D set of possible normals that satisfy the planarity and deflection constraints. Each of these normals lies on the pixel's refraction
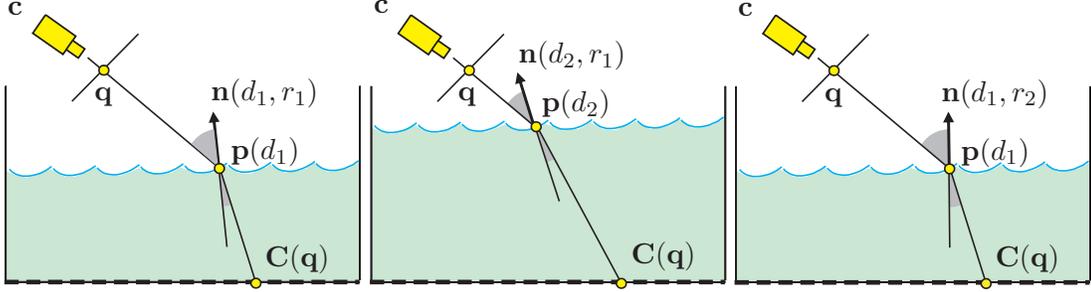
Figure 3.2: Single-viewpoint ambiguities. **Left:** A view of pixel **q**'s refraction plane is shown. **Middle:** Given a refractive index r_1, we can find, for each distance d_2 to the surface, a normal that refracts point $\mathbf{C}(\mathbf{q})$ to its corresponding pixel **q**. **Right:** Similarly, given a distance d_1, we can find, for each refractive index r_2, a normal that refracts point $\mathbf{C}(\mathbf{q})$ to pixel **q**.

plane and satisfies Eq. (3.1) for some value of the refractive index (Figure 3.2). Hence, the unit surface normal that satisfies Snell's law for pixel **q** can be expressed as a two-parameter family, $\mathbf{n}(d, r)$, parameterized by the distance d and the unknown refractive index, r. A closed-form expression for this normal as

$$\mathbf{n}(d, r) = r \, \| \mathbf{i}(d) \wedge \mathbf{o} \| \left(\frac{\mathbf{i}(d) - \big[\mathbf{i}(d) \cdot \mathbf{o}\big] \, \mathbf{o}}{\|\mathbf{i}(d) - \big[\mathbf{i}(d) \cdot \mathbf{o}\big] \, \mathbf{o}\|} \right) + \left(r \big[\mathbf{i}(d) \cdot \mathbf{o}\big] - 1 \right) \mathbf{o} \qquad (3.2)$$

where $\wedge$ denotes vector product; **o** is the direction of the ray through pixel **q**; and $\mathbf{i}(d)$ is the direction of the ray incident to the surface point $\mathbf{p}(d)$:

$$\mathbf{o} = \frac{\mathbf{c} - \mathbf{q}}{\|\mathbf{c} - \mathbf{q}\|}, \qquad (3.3)$$

$$\mathbf{i}(d) = \frac{\mathbf{c} - d \, \mathbf{o} - \mathbf{C}(\mathbf{q})}{\|\mathbf{c} - d \, \mathbf{o} - \mathbf{C}(\mathbf{q}))\|} \; . \qquad (3.4)$$

When the refractive index has a known value r_0, there is only one consistent normal, $\mathbf{n}(d, r_0)$, for each distance d. Sanderson *et al* [102] were the first to point out that this distance-

normal ambiguity for a pixel $\mathbf{q}$ can be resolved with the help of a second viewpoint.[1] Intuitively, a second viewpoint allows us to "verify" whether or not a particular distance hypothesis d is correct (Figure 3.1): given such a hypothesis and given the projection $\mathbf{q}'$ of point $\mathbf{p}(d)$ in the second camera, we simply need to verify that point $\mathbf{C}(\mathbf{q}')$ on the reference pattern refracts to pixel $\mathbf{q}'$.

While this hypothesis-verification procedure leads directly to an algorithm when the surface has a known refractive index, it leaves open the question of how to reconstruct surfaces whose 3D shape *and* refractive index are unknown. In this case, the surface normal lies in the full, two-parameter family, $\mathcal{N} = \{\mathbf{n}(d, r) \mid d, r \in \mathbb{R}^+\}$. One approach would be to use a third viewpoint to verify that a hypothetical refractive index r and distance d are consistent with the pixel-pattern correspondences in the three views.

Rather than use a third viewpoint, we prove that two views are, in fact, sufficient to estimate the 3D shape and refractive index of an unknown, generic surface. Intuitively, generic surfaces embody the notion of *non-degeneracy*—they are smooth surfaces whose differential properties remain unchanged if we deform their surface by an infinitesimal amount [66]. As such, they are especially suitable for modeling the complex, unconstrained shape of a liquid. Theorem 1 tells us if the liquid's surface is generic, the family, $\mathcal{N}$, of ambiguous solutions is discrete:

Theorem 1 $\mathcal{N}$ *is a zero-dimensional manifold for almost all pixels in the projection of a generic surface.*

Theorem 1 holds for continuous $\mathbf{C}(\mathbf{q})$ and suggests that it might be possible to compute the refractive index of a surface by choosing a single pixel $\mathbf{q}$ and finding the distance and refractive

---

[1] Sanderson *et al* [102] made this observation in the context of reconstructing opaque specular, rather than refractive, surfaces. Their analysis applies equally well to the case of refractive surfaces with a known refractive index.

index that are consistent with $\mathbf{C}(\mathbf{q})$ and the pixel-to-pattern correspondences in the second viewpoint[2]. In practice, image noise and the possibility of multiple discrete solutions dictate an alternative strategy, where measurements from multiple pixels contribute to the estimation of the refractive index. We consider the algorithmic implications of this result below.

3.3   Dynamic Refraction Stereo Algorithm

In order to reconstruct a liquid's surface at a time instant t, we need to answer three basic questions: 1) how to compute the pixel-to-pattern correspondence function $\mathbf{C}(\mathbf{q}, t)$, 2) how to compute the refractive index and 3) how to assign a distance and a normal to each pixel?

To compute $\mathbf{C}(\mathbf{q}, t)$ we rely on a procedure that computes the correspondences for time $t = 0$ and then propagates them through time using optical flow estimation.

Since the refractive index is the same for all pixels, we seek a value that most closely satisfies the refractive stereo geometry across all pixels and all frames. We perform a discrete 1D search in an interval of plausible refractive indices and, for each hypothetical value, attempt to reconstruct the scene for all pixels and frames. We then choose the value that produces the smallest reconstruction error. This leads to the following general algorithm, whose steps are discussed in the following sections.

Step 1 Initialize pixel-to-pattern correspondences, $\mathbf{C}(\mathbf{q}, 0)$.

Step 2 For each frame $t > 0$, estimate 2D optical flow to compute $\mathbf{C}(\mathbf{q}, t)$ from $\mathbf{C}(\mathbf{q}, t - 1)$.

---

[2]See Appendix A for the proof

Step 3 For every refractive index $r \in \{r_1, \ldots, r_n\}$, every frame t and every pixel $\mathbf{q}$,

- assuming refractive index r for the liquid, estimate the 3D position $\mathbf{p}$ and normal $\mathbf{n}$ of the surface point projecting to pixel $\mathbf{q}$ at time t

- estimate the reconstruction error (Section 3.4.2),

$$e(r,\, t,\, \mathbf{q}) = RE(\mathbf{p}, \mathbf{n}) \,. \tag{3.5}$$

Step 4 Set $r^* = \arg\min_r \sum_{t,\mathbf{q}} e(r,\, t,\, \mathbf{q})$ and return the distances and normals reconstructed with this index value.

Step 5 For each time t, fuse the pixel-wise 3D position and normal estimates to obtain a 3D surface.

3.4   Pixel-wise Shape Estimation

The key step in refraction stereo is an optimization procedure that assigns a 3D point $\mathbf{p}$ and a surface normal $\mathbf{n}$ to each pixel. The procedure assumes that the refractive index has a known value r and computes the $\mathbf{p}, \mathbf{n}$ that are most consistent with Snell's law and the pixel-to-pattern correspondence function for the input views.

For a given pixel $\mathbf{q}$, the optimization works in two stages. In the first stage, we conduct a 1D optimization along the ray through pixel $\mathbf{q}$. The goal is to find the distance d that globally minimizes a novel criterion, called the *refractive disparity (RD)*. This criterion is specifically designed to avoid instabilities due to degenerate refraction paths (e.g., when the liquid's surface is close to the reference pattern).

The optimal d-value gives us initial estimates, $\mathbf{p}(d)$ and $\mathbf{n}(d, r)$, for the 3D coordinates and surface normal of a point that projects to pixel $\mathbf{q}$. These estimates are further refined in a

second, bundle adjustment stage in which all five parameters (two for the normal, three for the position) are optimized simultaneously.

3.4.1 Measuring Refractive Disparity

Each value of d defines an implicit correspondence between four known points (Figure 3.1): pixel $\mathbf{q}$, point $\mathbf{C}(\mathbf{q})$ on the reference pattern that refracts to $\mathbf{q}$, the projection, $\mathbf{q}'$ of $\mathbf{p}(d)$ in the second viewpoint, and point $\mathbf{C}(\mathbf{q}')$. This correspondence must be consistent with Snell's law.

In their work on reconstructing mirror-like surfaces, Bonfort and Sturm [15] noted that such a correspondence gives us two "candidate" normals for $\mathbf{p}(d)$ which must be identical when this hypothesis is correct. These normals are obtained by applying Eq. (3.2) twice, once for each viewpoint. Specifically, the first normal, $\mathbf{n}_1 = \mathbf{n}(d, r)$, ensures that point $\mathbf{C}(\mathbf{q})$ on the reference pattern refracts to pixel $\mathbf{q}$ via point $\mathbf{p}(d)$. The second normal, $\mathbf{n}_2$, enforces a similar condition for the second viewpoint, i.e., it ensures that point $\mathbf{C}(\mathbf{q}')$ refracts to pixel $\mathbf{q}'$ via point $\mathbf{p}(d)$. We obtain $\mathbf{n}_2$ by applying Eq. (3.2) to pixel $\mathbf{q}'$, using its distance from point $\mathbf{p}(d)$. Since points on a smooth surface have a unique normal, a necessary condition for $\mathbf{p}(d)$ being on the "true" surface is that $\mathbf{n}_1 = \mathbf{n}_2$.

Unfortunately, even though it is possible, in principle, to directly measure the alignment of vectors $\mathbf{n}_1$ and $\mathbf{n}_2$, such a measurement becomes unstable when the distance between the surface and the reference pattern approaches zero. This is because as refraction diminishes, Eq. (3.2) becomes singular, normals cannot be estimated accurately, and the 3D reconstruction problem degenerates to standard stereo. In practice, this causes instability for low liquid heights, making direct comparison of normals uninformative and inappropriate for reconstruction.

Instead of measuring the alignment of the two normals $\mathbf{n}_1$ and $\mathbf{n}_2$ directly, we perform an
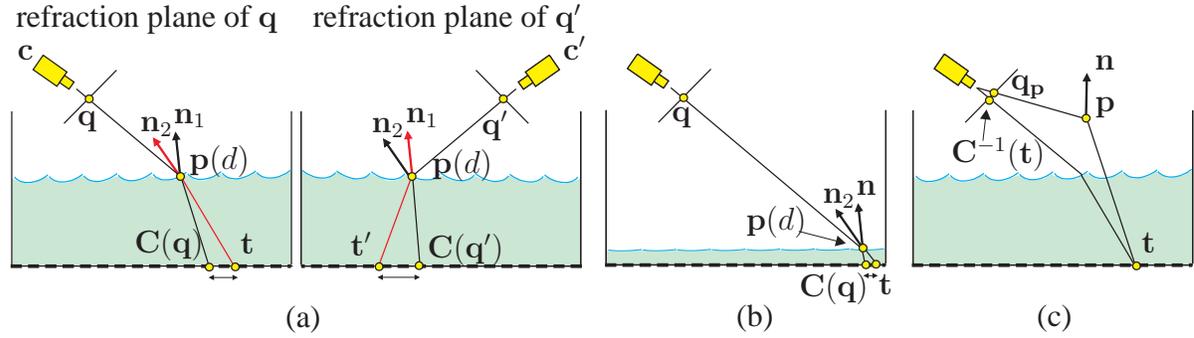
Figure 3.3: Optimization criteria for refraction stereo. (a) Measuring refractive disparity. Normals are drawn according to the refractions they produce. (b) For small surface-to-pattern distances, swapping n_1 and n_2 does not influence the distance between $\mathbf{C}(\mathbf{q})$ and $\mathbf{t}$ significantly. (c) Measuring image re-projection error at one of the viewpoints.

indirect measurement that is not singular when refraction diminishes. The main idea is that if $\mathbf{n}_1$ and $\mathbf{n}_2$ were truly aligned, "swapping" them would still force point $\mathbf{C}(\mathbf{q})$ to refract to pixel $\mathbf{q}$ and point $\mathbf{C}(\mathbf{q}')$ to pixel $\mathbf{q}'$. We therefore define the criterion by asking two questions (Figure 3.3a):

- suppose the normal at $\mathbf{p}(d)$ is $\mathbf{n}_2$; which point on the reference pattern will refract to $\mathbf{q}$?

- suppose the normal at $\mathbf{p}(d)$ is $\mathbf{n}_1$; which point on the reference pattern will refract to $\mathbf{q}'$?

Now suppose that points $\mathbf{t}, \mathbf{t}'$ are the points that refract to pixels $\mathbf{q}, \mathbf{q}'$, respectively. The distance between $\mathbf{t}$ and $\mathbf{C}(\mathbf{q})$ and, similarly, the distance between $\mathbf{t}'$ and $\mathbf{C}(\mathbf{q}')$, can be thought of as a measure of disparity. Intuitively, this distance tells us how swapping the normals $\mathbf{n}_1, \mathbf{n}_2$ affects consistency with the available pixel-to-pattern correspondences. To evaluate a hypothesis d we simply sum these distances:

Refractive Disparity

$$RD(d) = \|\mathbf{t} - \mathbf{C}(\mathbf{q})\|^2 + \|\mathbf{t}' - \mathbf{C}(\mathbf{q}')\|^2 . \tag{3.6}$$

When the distance between the true surface and the reference pattern is large, refractive disparity is equivalent to a direct measurement of the alignment between vectors $\mathbf{n}_1, \mathbf{n}_2$, i.e., it is zero if and only if $\mathbf{n}_1 = \mathbf{n}_2$. On the other hand, as the liquid's true surface approaches the reference pattern, refractive disparity diminishes. This is because the refractive effect of changing a point's surface orientation diminishes as well (Figure 3.3b). As a result, the minimization can be applied to any image pixel for which $\mathbf{C}(\mathbf{q})$ is known, regardless of whether or not the ray through the pixel actually intersects the liquid's surface.

To compute point $\mathbf{t}$ for a given d-value, we trace a ray from the first viewpoint through pixel $\mathbf{q}$, refract it at point $\mathbf{p}(d)$ according to normal $\mathbf{n}_2$, and intersect it with the (known) surface of the reference pattern. Point $\mathbf{t}'$ is computed in an identical manner. To find the distance d that globally minimizes refractive disparity along the ray we use Matlab's `fminbnd()` function, which is based on golden-section search [99].

3.4.2   Computing 3D Position and Orientation

Even though refractive disparity minimization yields good reconstructions in practice, it has two shortcomings. First, it treats the cameras asymmetrically because optimization occurs along the ray through one pixel. Second, it only optimizes the distance along that ray, not the 3D coordinates and orientation of a surface point. We therefore use an additional step that adjusts all shape parameters ($\mathbf{p}$ and $\mathbf{n}$) in order to minimize a symmetric image re-projection error.

To evaluate the consistency of $\mathbf{p}$ and $\mathbf{n}$, we check whether the refractions caused by such a point are consistent with the refractions observed in the input views. In particular, let $\mathbf{q}_\mathbf{p}, \mathbf{q}'_\mathbf{p}$ be the point's projections in the two cameras and suppose that $\mathbf{t}, \mathbf{t}'$ are the points on the reference pattern that refract to $\mathbf{q}_\mathbf{p}, \mathbf{q}'_\mathbf{p}$, respectively, via point $\mathbf{p}$. To compute the re-projection

error we measure the distance between pixels $\mathbf{q_p}, \mathbf{q'_p}$ and the "true" refracted image of $\mathbf{t}, \mathbf{t'}$ (Figure 3.3c):

$$RE(\mathbf{p}, \mathbf{n}) = \|\mathbf{q_p} - \mathbf{C}^{-1}(\mathbf{t})\|^2 + \|\mathbf{q'_p} - \mathbf{C}^{-1}(\mathbf{t'})\|^2 + \beta G(\|\mathbf{p} - \mathbf{p}_0\|; \sigma)^{-1} \qquad (3.7)$$

where $\mathbf{C}^{-1}(.)$ denotes the inverse of the pixel-to-pattern correspondence function, $G(.; \sigma)$ is the Gaussian with standard deviation σ, and $\mathbf{p}_0$ is the starting point of the optimization. The Gaussian term ensures that the optimization will return a point $\mathbf{p}$ whose projection always remains in the neighborhood of the originally-chosen pixel $\mathbf{q}$ in Steps 1 and 2 of the algorithm (Section 3.3). We used $\sigma = 4$ and $\beta = 200$ for all our experiments. To minimize the RE functional with respect to $\mathbf{p}$ and $\mathbf{n}$ we use the downhill simplex method [99].

3.5 Implementation Details

3.5.1 Estimating Pixel-to-Pattern Correspondences

Accurate 3D shape recovery requires knowing the pixel-to-pattern correspondence function $\mathbf{C}(\mathbf{q}, t)$ with high accuracy. While color-based techniques have been used to estimate this function for image-based rendering applications [21], they are not appropriate for reconstruction for several reasons. First, different liquids absorb different wavelengths by different amounts [116], altering a pattern's appearance in a liquid-dependent way. Second, since light absorption depends on distance traveled within the liquid and since this distance depends on the liquid's instantaneous shape, the appearance of the same point on a pattern will change through time. Third, the intensity of light transmitted through the surface depends on the Fresnel effect [34] and varies with wavelength and the angle of incidence. This makes it difficult to use color as a means to localize points on a pattern with near- or sub-pixel accuracy.

(a)　　　　　　　　　　　　　(b)　　　　　　　　　　　　　(c)
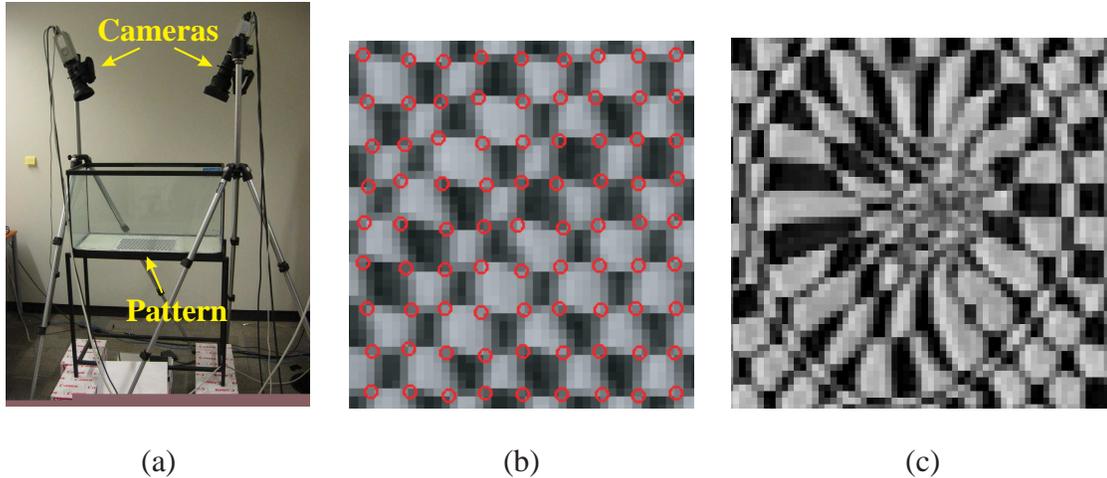
Figure 3.4: (a) Experimental setup. (b) Typical close-up view of pattern, seen through water surface. (c) Distorted view, corresponding to tracking failure at the central corners.

In order to avoid these complications, we use a monochrome checkered pattern and rely on corners to establish and maintain pixel-to-pattern correspondences (Figures 3.4a,b). We assume that the liquid's surface is undisturbed at time $t = 0$ and use the Harris corner detector [42] to detect corners at sub-pixel resolution. This gives us the initial pixel-to-pattern correspondences. To track the location of individual corners in subsequent frames while avoiding drift, we estimate flow between the current frame and the frame at time $t = 0$, using the flow estimates from the previous frame as an initial guess. We compute flow with a translation-only version of the Lucas-Kanade inverse-compositional algorithm [8] and use Levenberg-Marquardt minimization to obtain sub-pixel registration. This algorithm is applied to an 11×11 pixel neighborhood around each corner. We use the registration error returned by the algorithm as a means to detect failed localization attempts. In the case of failure, the flow computed for that corner is not used and the corner's previously known location is propagated instead. This allows our tracker to overcome temporary obscurations due to blur, splashes or extreme

refractive distortions (Figure 3.4c). The above procedure gives values of the correspondence function $\mathbf{C}(\mathbf{q}, t)$ for a subset of the pixels. To evaluate the function for every pixel, we use bilinear interpolation.

3.5.2   Fusing 3D Positions and Orientations

Refraction stereo yields a separate 3D position and 3D normal for each pixel. While this is a richer shape descriptor, the problem of reconstructing a single surface that is consistent with both types of data is still open. A key difficulty is that point and normal measurements have different noise properties and hence a surface computed via normal integration and a surface computed by fitting a mesh to the 3D points will not necessarily agree. As a first step, we used simulations and ground-truth experiments to estimate the reliability of each data source as a function of surface height, i.e., distance from the plane of the reference pattern (Figure 3.6). Since reconstructed normals are highly reliable for large heights, we used this analysis to set a height threshold below which normals are deemed less reliable than positions. That portion of the surface is reconstructed from positional data. For the remaining pixels, we reconstruct the surface via normal integration using the Ikeuchi-Horn algorithm [56] and merge the results. In cases where all reconstructed positions are above the height threshold, we rely on normal integration to compute 3D shape and use the average 3D position to eliminate the integrated surface's height ambiguity.

3.6 Experimental Results

3.6.1 Simulations

To evaluate the stability of our algorithms, we performed simulations that closely matched the experimental conditions in the lab (e.g., relative position of cameras, pattern-to-camera distances, feature localization errors, etc.). We simulated the reconstruction error for planar water surfaces as a function of the surface-to-pattern height, and for various levels of error in corner localization and camera calibration. We modeled the localization error by perturbing the image coordinates of the projected corners by a Gaussian with a fixed standard deviation. For each height, we reconstructed 10000 individual points and measured their deviation from the ground-truth plane (Figure 3.6). These simulations confirm that the accuracy of reconstructed normals degrades quickly for water heights less than $4mm$. Importantly, the accuracy of distance computations is not sensitive to variations in water height, confirming the stability of our optimization-based framework for refractive stereo (Section 3.4.1). We also compared the effect of localization error on the results (Figure 3.5).

In addition, we ran simulations to test the stability of refractive index estimation. We simulated a stationary sinusoidal surface whose average height was $40mm$ and whose amplitude was $2mm$. We then computed the total reconstruction error for various combinations of true and hypothesized refractive index values (Figure 3.7). We used a localization error of 0.1 pixels for these simulations, to reflect our actual experimental conditions. These simulations show that our objective function has a minimum very close to the expected refractive index. Note also that the valley around the minimum becomes more shallow as the refractive index increases.
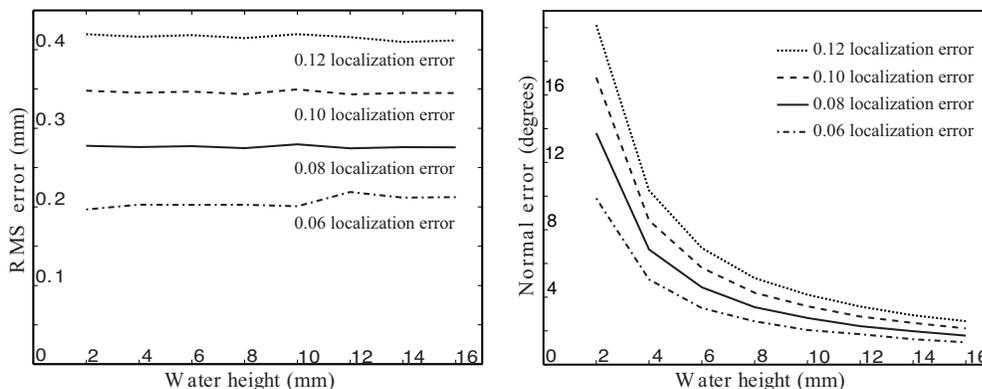
Figure 3.5: **Left:** Reconstruction accuracy as a function of water height, for the following values of pixel localization error from bottom to top $0.06, 0.08, 0.1, 0.12$. **Right:** Normal reconstruction error for the same pixel localization error values.

3.6.2 Experimental setup

Figure 3.4a shows our setup. The checkered pattern at the bottom of the tank was in direct contact with the water to avoid secondary refractions. During our experiments, the pattern was brightly lit from below to avoid specular reflections and to enable the use of a small aperture size for the cameras (and, hence, a large depth of field). Images were acquired at a rate of 60Hz with a pair of synchronized Sony DXC-9000 progressive-scan cameras, whose electronic shutter was set to $1/500sec$ to avoid motion blur. Both cameras were approximately 1 meter above the tank bottom and were calibrated using the Matlab Calibration Toolbox [17].

3.6.3 Accuracy experiments

Since ground truth was not available, we assessed our algorithm's accuracy by applying it to the reconstruction of flat water surfaces whose height from the tank bottom ranged from 4 to $15mm$. For each water height, we reconstructed a point $\mathbf{p}$ and a normal $\mathbf{n}$ independently for
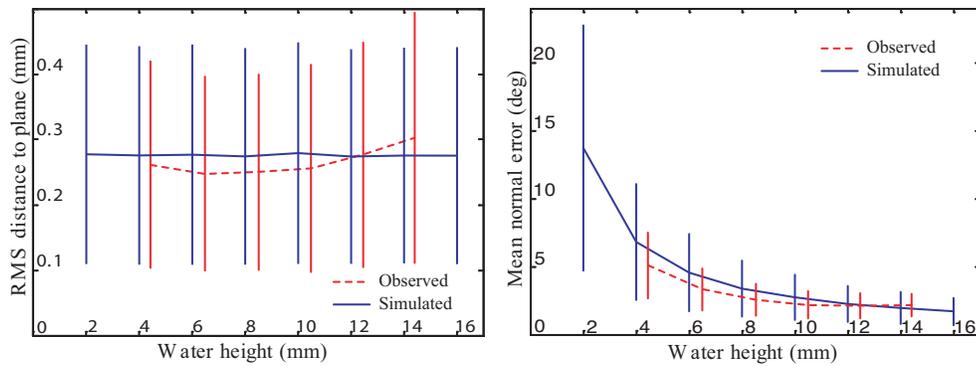
Figure 3.6: Reconstruction accuracy as a function of water height, for real (dotted line) and simulated (solid line) flat water surfaces. Bars indicate standard deviation. Simulations are for a 0.08-pixel localization error; in real flat water experiments, corner localization precision was measured to be approximately 0.1 pixels.

each of 1836 pixels in the two image planes, giving rise to as many 3D points and normals. No smoothing or post-processing was performed. To assess the accuracy of the reconstructed points, we fit a plane using least squares and measured the points' RMS distance from this plane. To assess accuracy in the reconstructed normals, we computed the average normal and measured the mean distance of each reconstructed normal from the average normal. These results, also shown in Figure 3.6, closely match the behavior predicted by our simulations.
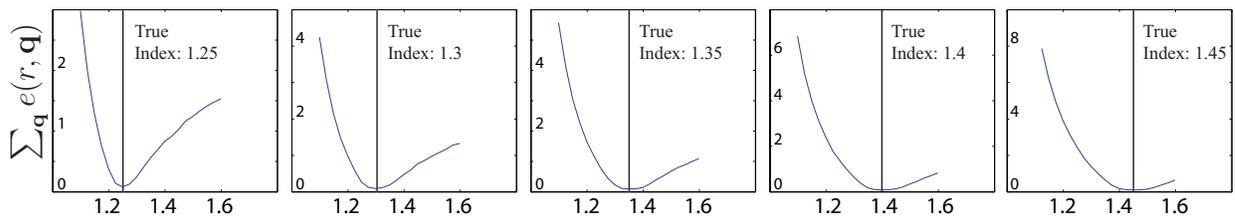
Figure 3.7: Total reconstruction error as a function of estimated refractive index for several indices increasing from left to right.

They suggest that reconstructions are highly precise, with distance variations around $0.25mm$, (i.e., within 99.97% of the surface-to-camera distance) and normal variations on the order of 2 degrees for water heights above $8mm$.

3.6.4   Experiments with dynamic surfaces

Figures 3.10, 3.11 and 3.12 show reconstructions for several dynamic water surfaces. The experiments test our algorithm's capabilities under a variety of conditions, from rapidly-fluctuating water that is high above the tank bottom, to water that is being poured in an empty tank, where the water height is very small and refraction is degenerate or near-degenerate for many pixels.

Several observations can be made from these experiments. First, our tracking-based framework allows us to maintain accurate pixel-to-pattern correspondences for 100s of frames, enabling dynamic reconstructions that last several seconds. Second, the reconstructed distances remain stable despite large variations in water height, and are accurate enough to show fine surface effects even in cases where the total water height never exceeds $6mm$ (e.g., the "pour" sequence). Third, the reconstructed normal maps, as predicted, show fine surface fluctuations at larger heights but degrade to noise levels for water heights near zero.

Qualitatively, when there is sufficient water in the tank, they appear to contain less noise than depth maps. Fourth, the normal integration algorithm that we are currently using seems to over-smooth fine surface details that are clearly present in the depth and normal maps. Hence, the question of how to best extract surfaces from the raw data provided by refraction stereo is still open. Our results suggest that a hybrid model that combines the benefits of the positional and normal data is necessary for accurate reconstruction such as [91]. In the following paragraphs, we present further analysis result sequences and their reconstructions [84].

Ripple sequence: The ripple sequence shows a drop of water dripping into a tank with

approximately $25mm$ of water in it. The drop causes layers of circular waves to emanate from the point of contact. There are a combination of large and fine scale waves in the sequence, making for an interesting reconstruction problem. We achieved a good reconstruction of depth and normals since the surface depth provided sufficient refraction for reliable readings in both categories. The normals provide better resolution of the fine ripples. The initial splash caused when the drop first hit the water produced strong distortions of the underlying pattern which our system was unable to track. We thus reinitialized the system immediately after the worst distortions had elapsed and were able to capture the expanding circular waves.

In the depth maps, at points of greatest distortion of the ripple, depth errors are more common, causing sharp peaks in the depth maps (bright or dark spots in Figure 3.10). Despite these depth errors, the corresponding normals appear to be correct and the normal maps are very smooth, allowing tiny leading ripples to be easily identified.

Pour sequence: This sequence shows water being poured into an empty tank, spreading across the pattern from left to right. This sequence tests our system's ability to cope with water heights of zero and close to zero. The sequence also exhibits large surface shape variation due to the ripples and several bubbles also formed on the surface during the sequence. Our reconstruction successfully handled the height variations using the refractive disparity. Our normal readings exhibited increased noise in the shallow portions, however we were able to reconstruct accurate depths (Figure 3.11). The large regions of noise in the normal maps correspond to areas with no water, however the heights are correctly recovered. Notice that the reconstructed height map is smooth at the water edges, showing robustness to shallow water.

The normal maps provide excellent fine detail of the water surface, lacking much of the noise evident in the height maps. Fine waves and ripples are visible in the normal maps in all three time instants of the pour sequence (Figure 3.11). Only the larger scale waves appear in
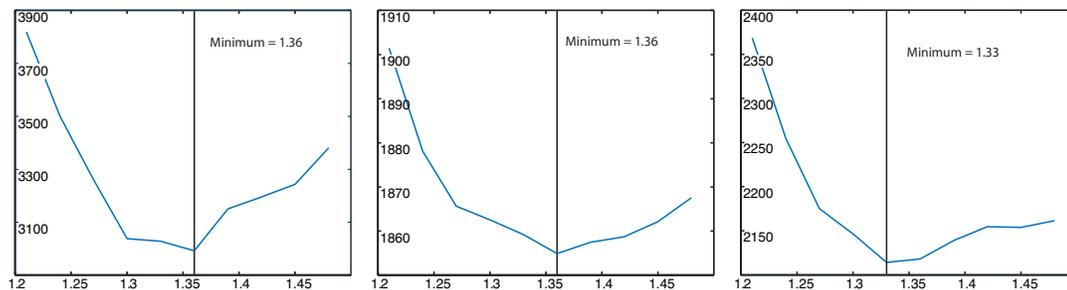
Figure 3.8: Total reconstruction error as a function of refractive index. **Left:** Combined error of several frames from the RIPPLE sequence. **Middle:** Combined error of several frames from the POUR sequence. **Right:** Combined error of several frames from the WAVES sequence.

the height map, where noise tends to obscure these details. Note that while our system is not designed to handle bubbles, the reconstruction actually captured indentations corresponding to the bubbles as well as the resulting ripples when they burst (see the middle of the normal map in the second time instant).

Waves sequence: This sequence shows waves propagating from the left to the right on a water surface at a height of approximately $24mm$. There are several interleaving wavefronts of various scales. This sequence provides the greatest variation in water height and overall roughness. Reconstructions from this dataset suffered from more calibration error and thus exhibit stronger noise. This is most noticeable in the moiré patterns appearing in both the depth and normal maps. Despite this, we did obtain interesting reconstructions of the rough water surface. We recover both the large and small scale wave fronts, which can be visually tracked across the sequence in the normal maps.

3.6.5 Refractive index determination

In addition to reconstructing both surface and normal data, our system was able to obtain an estimate of the refractive index of the transparent medium. We ran our algorithm on all three sequences of Figures 3.10-3.12, over a range of seven frames for each. Figure 3.8 shows the total reconstruction error corresponding to specific values of the refractive index, for the each dataset. The curves exhibit a minimum near the correct refractive index for water, 1.33, confirming the predictions of Theorem 1. This is because Step 4 of our algorithm enforces a very strong *global* constraint: the light path of every pixel at every time instant and at every viewpoint must be consistent with the *same* refractive index value.

3.7 Discussion

3.7.1 Ambiguous surfaces

In refractive index determination we must examine the possibility that two surfaces with different refractive indices produce the same observed stereo image pair. Our initial analysis, described below, indicates that pairs of surfaces with different refractive indices that produce the same images do exist, but they do not have a "simple" shape: indeed, the points on a pair of ambiguous surfaces must satisfy a joint system of very complex trigonometric equations for such an ambiguity to exist.

We use a numerical approach to construct point samples on such surface pairs as follows: given the refractive index r of the true surface S and the refractive index r_a of an ambiguous surface R, we construct points and normals that lie on both S and R. First we begin with a seed point $\mathbf{p}_1$ on S which is imaged at $\mathbf{q}_1$ in $\mathbf{c}$ (Figure 3.9). Then we find a second point $\mathbf{a}_1$

(a)   (b)
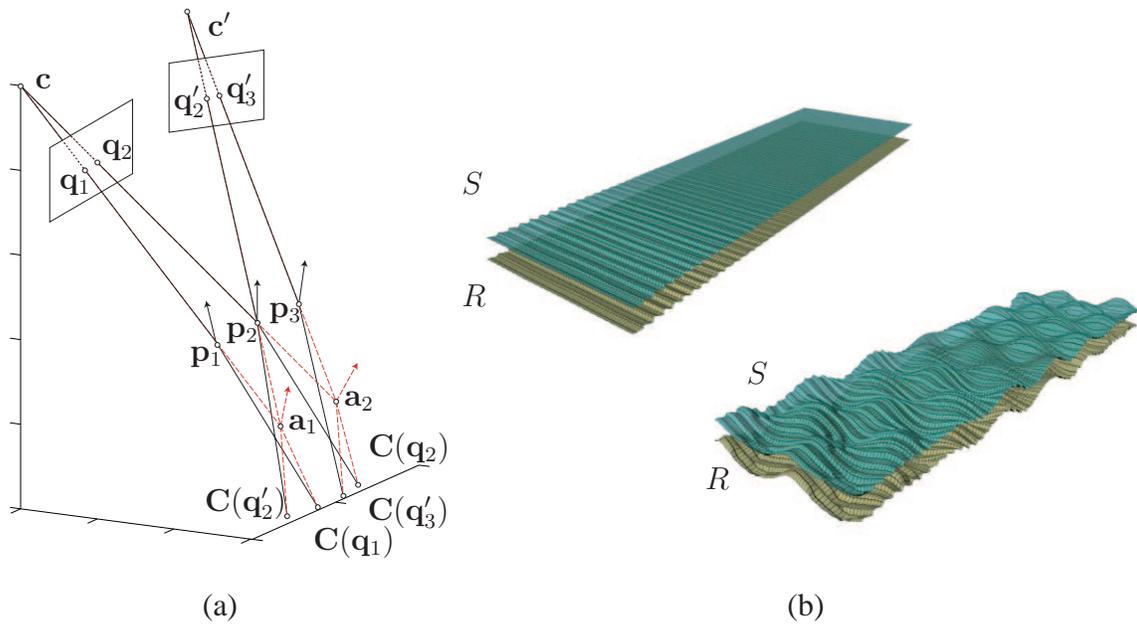
Figure 3.9: (a) Ambiguous surface construction: Points $\mathbf{p}_1$, $\mathbf{p}_2$ and $\mathbf{p}_3$ lie on surface S, points $\mathbf{a}_1$ and $\mathbf{a}_2$ lie on the ambiguous surface R. Note that correspondence functions $\mathbf{C}(\mathbf{q}'_2)$, $\mathbf{C}(\mathbf{q}'_3)$, $\mathbf{C}(\mathbf{q}_1)$ and $\mathbf{C}(\mathbf{q}_2)$ all agree with both the true and ambiguous surface points and their normals. (b) Ambiguous surface pairs constructed according to Section 3.7.1. **Left**: The depths of S were chosen to fit a plane. **Right**: The depths of S were chosen to fit a sinusoidal surface.

that lies along the ray from $\mathbf{q}_1$ to $\mathbf{p}_1$ and on the ambiguous surface R. The depth of $\mathbf{a}_1$ and its normal $\mathbf{m}_1$ must be chosen such that the pixel ray refracts to the same location on the tank bottom described by $\mathbf{C}(\mathbf{q}_1)$. Next, we project $\mathbf{a}_1$ into view $\mathbf{c}'$ and intersect this ray with S to obtain another point $\mathbf{p}_2$. The normal of $\mathbf{p_2}$ is constrained by the condition that it refract the pixel ray from $\mathbf{c}'$ must refract to $\mathbf{C}(\mathbf{q}'_1)$. From our initial seed point, this process gives us two new points: $\mathbf{a}_1$ on R and $\mathbf{p}_2$ on S. We can repeat this process given additional new seed points or by extending the construction by repeating the above steps with $\mathbf{p}_2$ as the seed. Figure 3.9(a) shows this construction repeated twice. Since we have one degree of freedom in

the depth/normal of $\mathbf{a}_1$, this means there is a family of surfaces that are ambiguous, yet highly constrained. For instance, if we attempt to create S and R as parallel planes, the normals associated with the points cannot satisfy these constraints and agree with the global planar normal. Figure 3.9(b) shows two surface pairs created from a sparse set of points as described. Intermediate surface points were interpolated according to the sparse points and their normals. These examples show global fitting of a desired shape (e.g. plane or sinusoidal surface) but locally it does not match since the ambiguity must be maintained by specific curvature at these sparse points. If we were to construct these surfaces as shown and then test for ambiguity at the intermediate points, the surfaces would not be ambiguous

So, while ambiguities exist they are not as important in practice because in our context the scene is dynamic: in our algorithm, a single refractive index value must account for the refractions produced by the *entire sequence* of 3D surfaces of the liquid, not just the 3D surface in a single instant. In fact, refractive index estimation exploits the dynamic/statistical nature of liquids in three ways: 1) the surface is highly variable and hence we observe many different, complex surfaces with the *same* refractive index during image acquisition, 2) their deforming surface is unlikely to globally match one of the "special," ambiguous shapes and 3) even if it does, it is unlikely that such a "special" shape will occur for many time instants. In practice, this allows us to side-step the issue of ambiguities by enforcing refractive index consistency with all available data (multiple time instants, even multiple acquisition experiments) using our search-based refractive index estimation algorithm. Experimentally, the lack of shape-index ambiguities over an acquired dataset is confirmed by error curves that have only one (global) minimum, as in Figure 3.8.

3.7.2   Pixel-to-pattern function

In the description of the algorithm and implementation, we assumed that $\mathbf{C}()$ was invertible. We note, however, that $\mathbf{C}()$ can be many-to-one and not generally invertible. The conditions that cause $\mathbf{C}()$ to be invertible were noted by Murase, deemed reasonable, and used in his work on liquid reconstruction. From a technical standpoint, however, our analysis does not require $\mathbf{C}()$ to be globally invertible: all we need is that it is *locally* invertible, i.e., for almost all pixels $\mathbf{q}$ (in a measure-theoretic sense), the restriction of $\mathbf{C}()$ to some open neighborhood of $\mathbf{q}$ is an invertible function. This does permit the occurrence of isolated singularities (i.e., pixels or image curves where $\mathbf{C}()$ is not invertible for any neighborhood).

For example, if we were to take a simple 2D scene such as a surface defined by $y = cos(x)$ with the camera looking down the $-y$ axis, $\mathbf{C}()$ is not globally invertible. The scene is, however, locally invertible for all values of x except two: for a given refractive index value and a camera located at infinity, there are only two incoming rays/pixels where local invertibility breaks down: these rays hit the $cos(x)$ curve near its inflection points, where the mapping from incoming rays to points on the x-axis "folds" onto itself. More generally, the singularities where local invertibility breaks down have properties analogous to the singularities of the Gauss map where, generically, the mapping from surface points to their normals is singular either on parabolic curves, corresponding to "folds" of the Gauss map, or on isolated points, corresponding to "cusps" of the map.

We also examine in further detail, how the flow propagates from $\mathbf{C}(t-1)$ to $\mathbf{C}(t)$. There are two cases: (1) The Lucas-Kanade algorithm is able to localize in frame t a corner that was also localized in frame $t-1$. In this case, the flow vector assigned to the corner at time $t-1$ is its displacement between the two frames. This process is completely local and is well defined wherever $\mathbf{C}()$ is locally- (but perhaps not globally-) invertible. (2) The LK algorithm

fails to localize the corner at frame t. This does *not* cause a breakdown of flow estimation for subsequent frames. In this case, the algorithm interpolates the flow vectors computed at four neighboring corners in frame $t-1$ in order to assign a flow vector to the corner that was lost in frame t. Bilinear interpolation is used, with weights determined by the corners' distance from each other. We then use the position in frame $t-1$ of the lost corner and the interpolated flow vector to assign it a "virtual" position in frame t. This position is used to initialize the LK algorithm in frame $t+1$, in an attempt to re-localize the lost corner. In case of failure at $t+1$, propagation is repeated until the corner is re-acquired. Since these distortions are local and persist for just few frames, we have found that the strategy works well in practice and has enabled propagation of pixel-to-pattern correspondences for $100s$ of frames.

Our current implementation does not check for the possibility that after a tracking failure (i.e., singularity of $\mathbf{C}()$) a corner at frame $t-1$ appears in more than one location in frame t (or vice-versa). While is certainly possible to do so, the sequences we have acquired suggest that such events are very transient and cause significant distortions in the local neighborhood of a corner, making it very hard to localize it, let alone identify multiple images of it. In such cases, flow propagation allows the corner to be re-acquired when distortions are reduced.

3.8   Concluding Remarks

Liquids can generate extremely complex surface phenomena, including breaking waves, bubbles, and extreme surface distortions. While our refraction stereo results are promising, they are just an initial attempt to model liquid flow in relatively simple cases. Our ongoing work includes (1) reconstructing surfaces that produce multiple refractions [69], (2) reconstructing liquids by exploiting their refractive and reflective properties (e.g., by also treating them as

mirrors), and (3) "reusing" captured 3D data to create new, realistic fluid simulations.
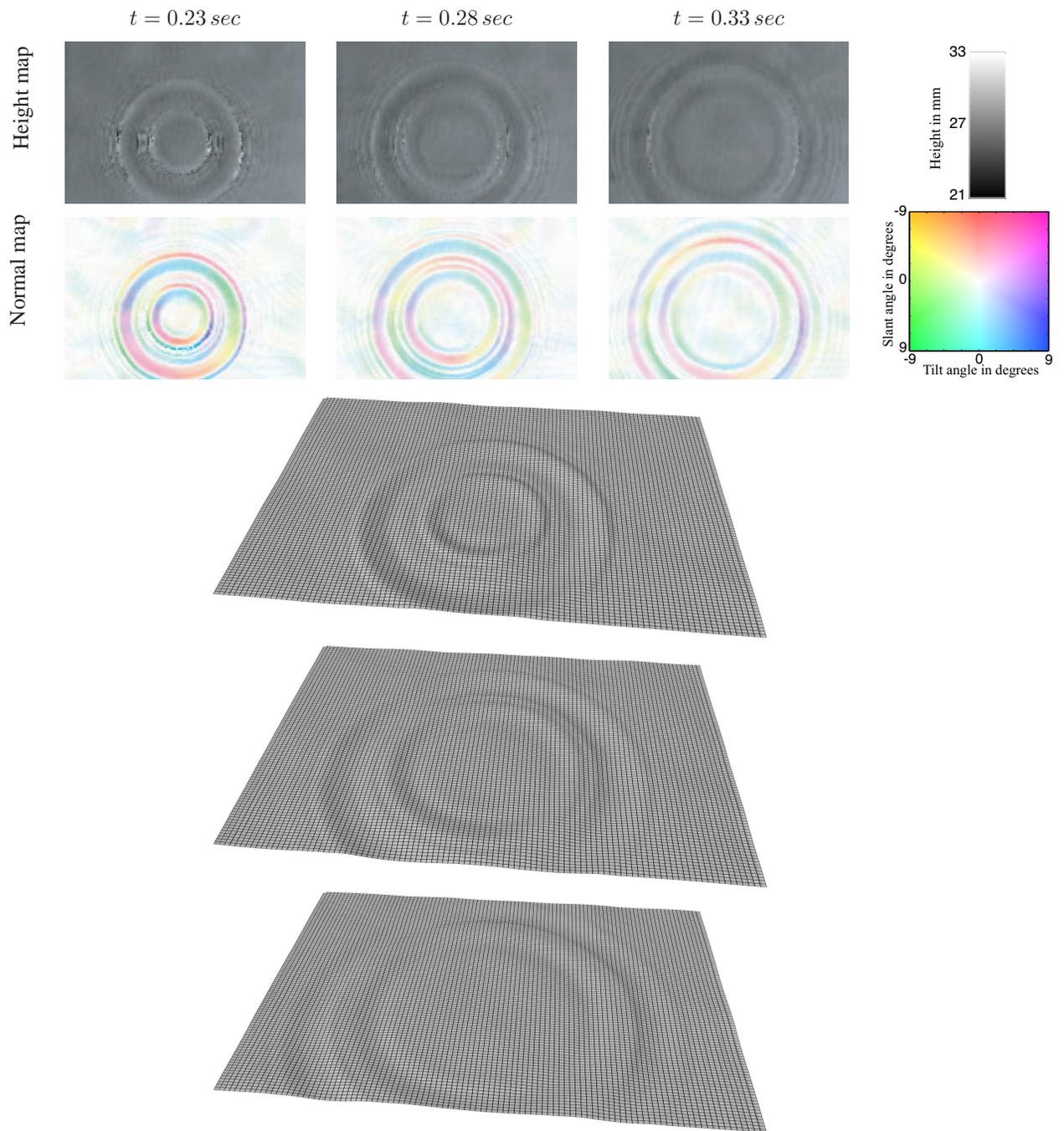
Figure 3.10: RIPPLE Sequence. All maps correspond to a top view of the tank and show raw, per-pixel data. The mesh images show a surface fit to the data scaled by 5 in the vertical axis.
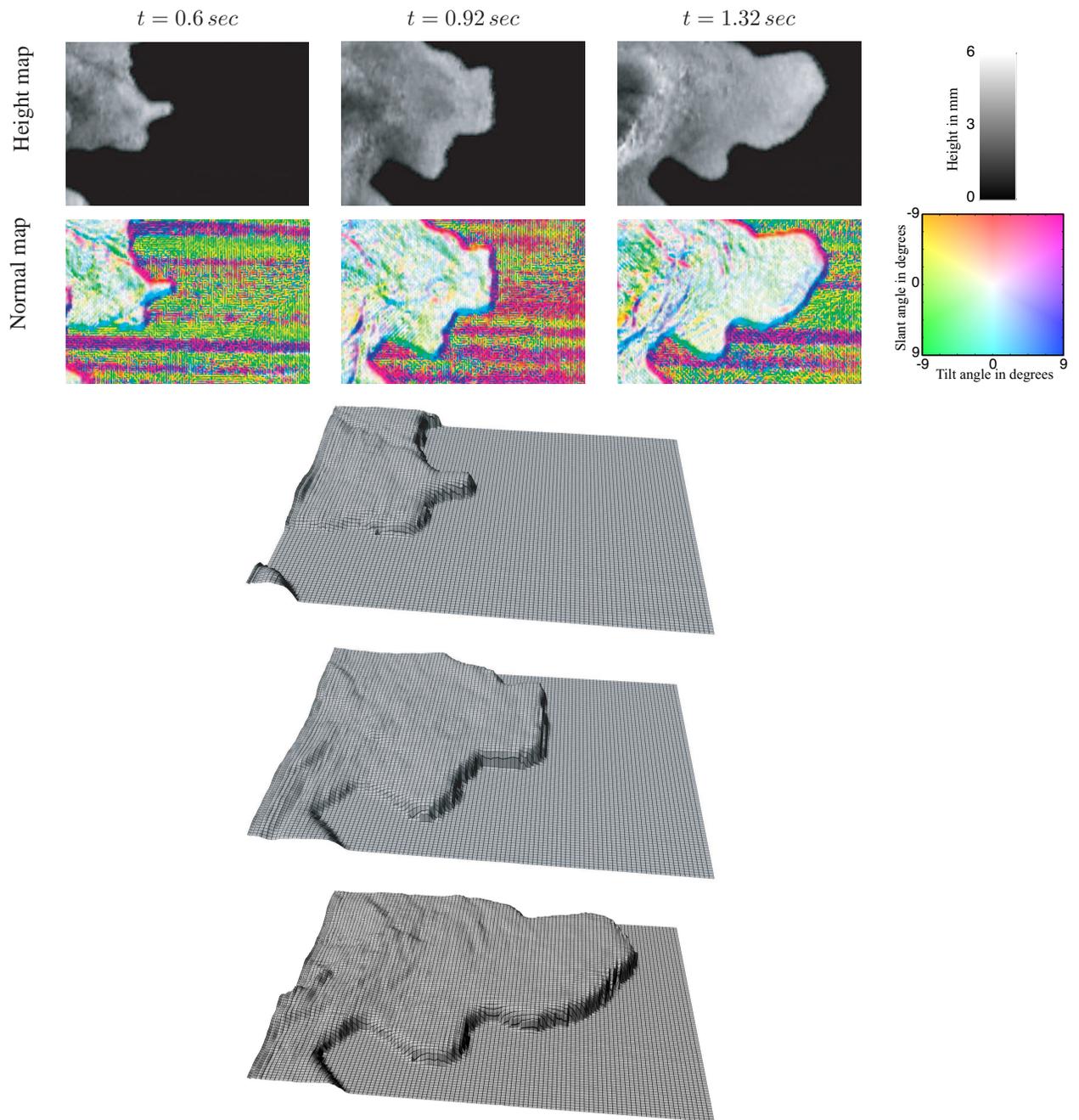
Figure 3.11: POUR Sequence. All maps correspond to a top view of the tank and show raw, per-pixel data. The mesh images show a surface fit to the data scaled by 5 in the vertical axis.
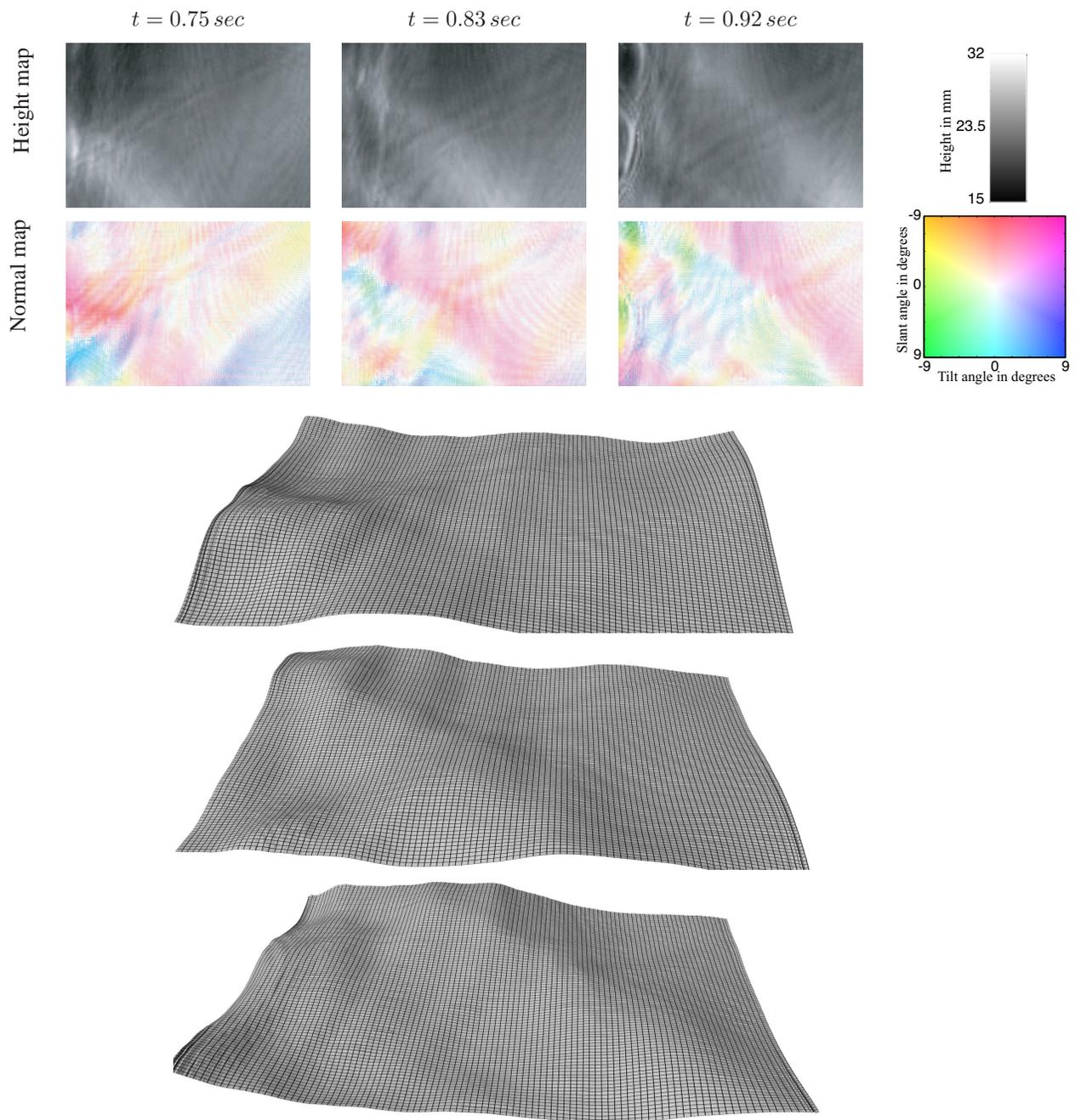
Figure 3.12: WAVES Sequence. All maps correspond to a top view of the tank and show raw, per-pixel data. The mesh images show a surface fit to the data scaled by 5 in the vertical axis.

Chapter 4

Scatter-Trace Photography

"Don't tell me the moon is shining; show me the glint of light on broken glass."

-Anton Chekhov

4.1   Introduction

A major ingredient in the success of recent 3D photography algorithms is their ability to deal with *surface inhomogeneity*, i.e., to produce accurate 3D models even when a scene's surfaces span a broad range of shape and material properties [24, 67, 141]. These algorithms apply exclusively to opaque surfaces that scatter incident light, and cannot handle scenes that contain transparent or highly-reflective media. For such scenes, the state of the art in reconstruction [9, 69, 83, 84, 111, 123, 137] is still confined to the simplest possible case—a surface bounding a single, homogeneous, transparent volume with no internal structures and no occlusion—and even this case cannot be solved without further assumptions (e.g., partially-known geometry [9, 83], a volume that causes no more than two refractions [69], or ability to

61

Figure 4.1: Three objects used in our experiments: A hand-made solid crystal sculpture with a painted interior (about 30cm tall), a partly-full juice bottle, and a decorative glass bottle with an opaque hand-painted interior. The red arrow highlights the pixel used in the example of Figure 4.16, bottom right.

immerse in a refractive-index-matched liquid [123]). Unfortunately, while objects with transparent media are very common (Figure 4.1), they rarely appear in isolation and rarely have a simple enough shape to fall within the realm of existing techniques. For such objects, research has concentrated on capturing their appearance rather than reconstructing them [82, 143].

The difficulty in reconstructing such scenes stems from the complex relation between their appearance, their exterior 3D shape, and the structure of their interior. This appearance can be heavily influenced by several light transport phenomena, including one or more refractions; total internal reflection; absorption and scattering at an interior interface; and reflection at an exterior surface. Inverting the interior light transport process under these conditions has proved very difficult.

Motivated by these difficulties, this paper develops an approach for reconstructing the exterior of general, inhomogeneous transparent scenes with three basic goals in mind:

- **Invariance to scene interior:** To the extent possible, reconstruction performance should depend on the scene's exterior surfaces, not the structure and complexity of media in the interior.

- **Robustness to spatially-varying reflectance and transmittance:** Reconstruction algo-
  rithms should be able to handle a wide range of surface reflectance and transmittance
  properties.

- **Compatibility with existing methods:** It should be possible to leverage developments
  in 3D photography of opaque scenes to treat issues such as noise, missing data, and
  occlusions.

To achieve these goals, we rely on the well-known fact that transparent scenes reflect some of
the incident light, thereby behaving as partial, non-ideal specular reflectors [40, 117]. Using
this as a starting point, we develop a novel technique based on *scatter-trace photography* that
is specifically designed to analyze these reflections.

Scatter-trace photography involves capturing images of the scene from one or more view-
points while moving a proximal light source to a 2D (or 3D) set of positions. This produces a
2D (or 3D) set of measurements per pixel, which we call the pixel's *scatter trace.* Intuitively,
the scatter trace of a pixel can be thought of as a "photograph" of the trajectories that light
followed before interacting with the scene, and before arriving at the given pixel (Figure 4.2).

The key property of the scatter trace is that direct surface reflection leaves a highly-constrained
geometric "signature" in it, even when light transport within the scene's interior is exceedingly
complex. Moreover, this signature is especially prominent when the direct reflection compo-
nent includes a non-negligible contribution from specular reflection. This observation leads
to three main results. First, it gives rise to a geometry-based method for enhancing the con-
trast of the direct reflection component in each scatter trace, relative to all other modes of
light transport. Second, it allows us to reduce reconstruction of inhomogeneous scenes with
non-negligible specular reflectance to a generalized form of stereo matching, where we es-

tablish correspondences by comparing appropriately-processed scatter traces rather than raw pixel intensities. Third, we show that this process provides detailed information about surface orientation, at sub-pixel resolution.

Our work relies on the existence of a non-negligible specular reflection component to recover 3D shape and, as such, it is closely related to *specular stereo* methods [15, 102, 103, 120]. These methods recover shape by analyzing the distorted appearance of patterns placed near an opaque, mirror-like scene. Of particular relevance is the work of Bonfort, Sturm and Gargallo [16] and Kutulakos and Steger [69], whose goal is to reconstruct the light path that connects each pixel in the image with two known 3D points that project to it. Both approaches rely on an idealized image formation model, where light is transported along an infinitely-thin, single-bounce path corresponding to direct specular reflection off a mirror. Our approach can be thought of as generalizing these methods to the case of non-ideal inhomogeneous scenes, i.e., scenes whose interior contributes significantly to appearance and whose exterior is not perfectly specular.

Although we do not explicitly decompose photos into direct, indirect, and specular components, a weaker form of separation—relative contrast enhancement—occurs implicitly as part of our generalized stereo matching procedure. In this respect, our work can be viewed as a geometry-driven alternative to existing layer decomposition methods (direct vs. indirect [89, 110], specular vs. diffuse [25], reflected vs. transmitted [94, 117]). These methods employ a variety of tools, including active illumination [89, 110], polarization analysis [25], optical flow estimation [117], and natural image statistics [73]. Unfortunately, none of them apply to the case of general, inhomogeneous transparent scenes. For example, recent algorithms for direct/indirect separation [89, 110] break down in the presence of strong specular reflection, and polarization state stops being a robust separation cue in the presence of refrac-

tion, reflection and scattering in a scene's unknown interior.

Our work offers four main contributions over the current state of the art. First, we derive a simple "scatter-trace stereo" algorithm for reconstructing the exteriors of scenes with transparent surfaces and inhomogeneous interiors. Second, we show that scatter-trace photography provides a natural means for revealing the scattering properties of complex scenes. Third, by reducing the reconstruction problem to a simple pairwise pixel-matching criterion, our work suggests that reconstruction of inhomogeneous scenes is possible by simply replacing the "data term" in existing stereo formulations. Fourth, our results show that scatter trace analysis enables reconstruction in the presence of complex shapes and spatially-varying surface reflectance and transmittance properties. We are not aware of other image-based methods capable of reconstructing scenes of this optical complexity.

4.2 Scatter-Trace Photography

Scatter-trace photography provides a convenient way to capture the interaction of a scene with proximal point light sources and viewpoints. The most general way of "probing" this interaction is to place a point light source at some position near the scene, emit radiance only within a differential solid angle along some direction, and then measure incident radiance at some other position and direction (Figure 4.3a). The set of all such measurements is a ten-dimensional function that we call the *plenoptic scatter function*. This function describes how the scene scatters incident light and takes into account distance-dependent effects (e.g., that objects appear dimmer as the point light source moves farther away from them). As such, the plenoptic scatter function generalizes the familiar notions of the 4D light field [74], the 5D plenoptic function [2], and the 8D reflectance field [25, 143].
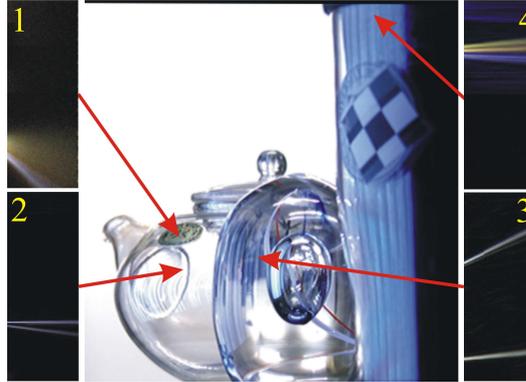
Figure 4.2: *Center, back to front*: a teapot made of thin glass about 2mm-thick; a solid crystal ornament with an internal air bubble; a mug with a 2mm-thick enclosure made of clear plastic and an opaque interior cavity made of a purple specular material (see Figure 4.16, bottom right for another view). *Sidebars:* Scatter traces for the four pixels indicated by arrows. These pixels receive light (1) by direct reflection at an opaque point; (2) by direct reflection and by internal reflection at the back-surface of the thin teapot wall; (3) by direct reflection and by a secondary internal reflection; and (4) by direct reflection, internal reflection at the back-surface of the clear enclosure, and internal reflection off the purple interior. Note the distinct "traces" associated with each propagation mode.

The plenoptic scatter function has extremely high dimensionality and would be very difficult to capture. Scatter-trace photography captures a specific 5D (or 4D) "slice" of this function that is both easy to capture and provides strong information about scene geometry. In particular, suppose that we observe the scene from a single viewpoint while illuminating it with an isotropic point light source (Figure 4.3b). In this case, every position (x, y, z) of the light source produces a distinct image. This image represents radiance that leaves the source equally in all directions and reaches the camera's image plane after interacting with the scene.[1]

Scatter-trace photography involves moving such a light source to every position within a

---

[1]To simplify our discussion, we assume here that image pixels have been calibrated to measure radiance directly [26].
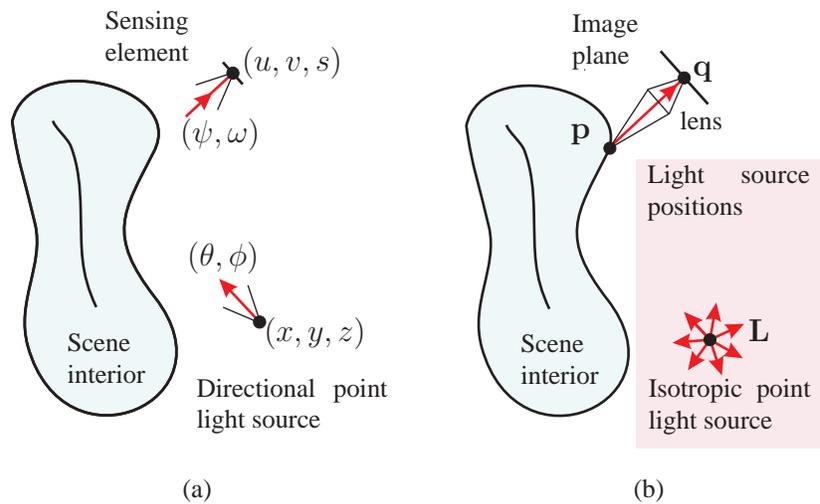
Figure 4.3: (a) To obtain a sample of the plenoptic scatter function we must choose a position and orientation for both the light source and the sensor (five degrees of freedom each). (b) To measure the scatter trace of pixels from a fixed viewpoint, we move the light source to every point inside a region of space (shaded).

volume of space, to obtain a 3D set of 2D images. Every pixel $\mathbf{q}$ on the camera's image plane is then associated with a volume of measurements, one for each light source position. We call these measurements the scatter trace of pixel $\mathbf{q}$:

Definition 1 (Scatter Trace of pixel q) $T_{\mathbf{q}}(\mathbf{L})$ *is the incident radiance at* $\mathbf{q}$ *when the light source is at point* $\mathbf{L}$.

Note that if a pixel's scatter trace is zero for some light source position, then no light passing through that position can possibly contribute radiance to that pixel. Therefore, the non-zero region of a pixel's scatter trace is the set of all points that light can pass through to reach that pixel.
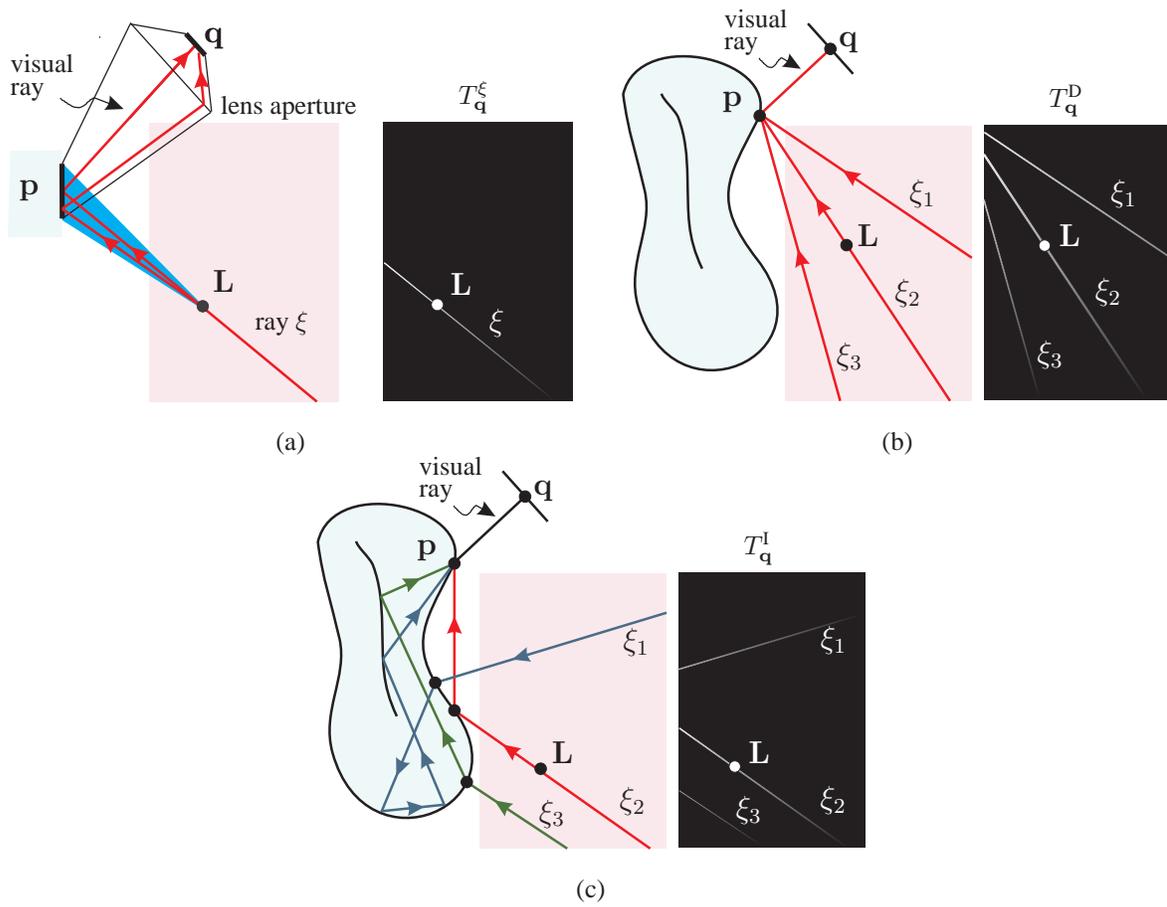
Figure 4.4: (a) Scatter trace due to reflection off a planar mirror. The radiance contributing to a finite-sized pixel $\mathbf{q}$ is transported along a bundle of rays that originate at $\mathbf{q}$; pass through the lens aperture; are specularly reflected at a small in-focus planar facet $\mathbf{p}$; and converge at the light source position. The radiance at pixel $\mathbf{q}$ is therefore bounded from above by the total radiance incident at $\mathbf{p}$, i.e., the radiance transported along the blue-shaded region. In the limit, as the size of $\mathbf{q}$, $\mathbf{p}$ and the aperture goes to zero, the scatter trace becomes concentrated on a single ray ξ and defines the *impulse scatter trace* along ξ. (b) Generalized specular reflection produces a "fan" of impulse scatter traces (only 3 are shown, for clarity). (c) The indirect scatter trace produced by light transport along 3 paths ending at $\mathbf{q}$.

4.2.1 Direct and Indirect Scatter Traces

Since light transport is linear, we can express the scatter trace as a sum of two components

$$T_{\mathbf{q}} = T_{\mathbf{q}}^{\mathrm{D}} + T_{\mathbf{q}}^{\mathrm{I}} \,, \tag{4.1}$$

where $T_{\mathbf{q}}^{\mathrm{D}}$ represents the contribution of direct surface reflection and $T_{\mathbf{q}}^{\mathrm{I}}$ represents the contribution of indirect light transport (i.e., refraction, total internal reflection, surface inter-reflection, caustics, etc).

In general, the indirect component will be a significant fraction of the total scatter trace, and this fraction will vary from pixel to pixel (Figure 4.2). Fortunately, the direct and the indirect components of the scatter trace have a distinct spatial structure. Here we exploit this difference to enhance the contrast of the direct component and use it for single-view reconstruction (Section 4.3.2) and multi-view stereo matching (Section 4.4).

Since our analysis relies on the spatial structure of the scatter trace, we consider below the scatter trace produced by three basic types of light transport.

Direct reflection without scattering Consider an infinitesimally-small pixel $\mathbf{q}$ that is perfectly focused at a point on a planar mirror. As the lens aperture shrinks to a zero, only light sources along one incident ray ξ will contribute to the pixel's radiance (Figure 4.4a). This ray is along the direction of specular reflection. We call the resulting scatter trace the *impulse scatter trace*, $T_{\mathbf{q}}^{\xi}$, of pixel $\mathbf{q}$.

For a given light source position along this ray, the radiance received at the in-focus surface point obeys a squared-distance falloff [50]. Since the radiance at $\mathbf{q}$ cannot be larger than this radiance, the impulse scatter trace is a single "streak," whose intensity diminishes with distance from the point of reflection.

Generalized specular reflection If the surface point projecting to an infinitesimally-small pixel $\mathbf{q}$ is not a planar mirror, the pixel will receive light even from light sources that are not along a single ray. To account for this behavior, we model surface micro-geometry as a distribution of planar micro-facets that vary arbitrarily from point to point on the surface [29]. Furthermore, we assume that each micro-facet acts as a mirror reflector, with minimal masking or shadowing [88]. This model accurately represents the reflectance properties of smooth surfaces (e.g., glass, polished metal) and accounts for high-curvature reflectors.

Consider the surface point $\mathbf{p}$ that projects to pixel $\mathbf{q}$. We can express the point's micro-facet distribution as a probability distribution $D(\theta, \phi)$ over the unit sphere, with the angles (θ, ϕ) corresponding to a unique normal. In this case, the scatter trace is a weighted superposition of impulse scatter traces, one for each incoming ray (Figure 4.4b):

$$T_{\mathbf{q}}^{\mathrm{D}}(\mathbf{L}) = \int_{\text{rays through } \mathbf{p}} D(\theta^\xi, \phi^\xi) \, T_{\mathbf{q}}^\xi(\mathbf{L}) \, d\xi \,, \tag{4.2}$$

where the normal (θ^ξ, ϕ^ξ) is the bisector of ray ξ and the visual ray through pixel $\mathbf{q}$.

Intuitively, as the surface at the point departs from a planar mirror, the scatter trace spreads into a "fan" of converging streaks (Figure 4.2, pixel 1). The point of convergence of this fan is always the surface point projecting to pixel $\mathbf{q}$. Moreover, the fan's intensity decreases monotonically in a radial direction away from that point.

General indirect reflection/transmission Now suppose that all light received at pixel $\mathbf{q}$ is due to one or more indirect reflection and/or transmission events. Again, we can express the scatter trace as a weighted combination of impulse scatter traces, although the set of scatter traces participating in this combination is much more general: each impulse scatter trace in this set represents the contribution of light that travels along an arbitrary ray ξ until it hits the

object, and then follows a general piecewise-linear path to pixel $\mathbf{q}$ (Figure 4.4c):

$$T_{\mathbf{q}}^{\mathrm{I}}(\mathbf{L}) = \int_{\text{all rays}} \tau_{\mathbf{q}}(\xi)\, T_{\mathbf{q}}^{\xi}(\mathbf{L})\, d\xi \;. \tag{4.3}$$

Here, the weight $\tau_{\mathbf{q}}(\xi)$ is the fraction of radiance transported to pixel $\mathbf{q}$ from the point of first contact of ray ξ with the object.

Unlike the case of direct reflection, the scatter trace produced by indirect reflection is much less constrained (Figure 4.2, pixels 2-4). In general, the "streaks" it contains will *not* converge to a single point and, even if they do (which is a non-generic event), their point of convergence is not constrained to lie on the visual ray through pixel $\mathbf{q}$.

2D scatter traces So far, we have assumed that the point light source moves to a 3D set of positions. This is rather inefficient. In practice, we obtain an equivalent set of measurements by illuminating the scene with a *linear* light source (e.g., aligned with the z-axis) and moving it to a 2D set of positions (e.g., on the xy-plane). This procedure gives us a reduced, 2D scatter trace per pixel that has exactly the same properties as its 3D counterpart (see Appendix B). Our analysis below applies both to 3D and 2D scatter traces.

4.3   3D Shape from Scatter-Trace Constraints

The previous section showed that the direct component of a pixel's scatter trace is a superposition of impulse scatter traces (i.e., "streaks") that satisfy three basic constraints:

- **Viewing-ray intersection:** They must all intersect the pixel's viewing ray.

- **Convergence:** They must converge to a single point on the viewing ray, and this point must coincide with the surface point that caused the direct reflection.
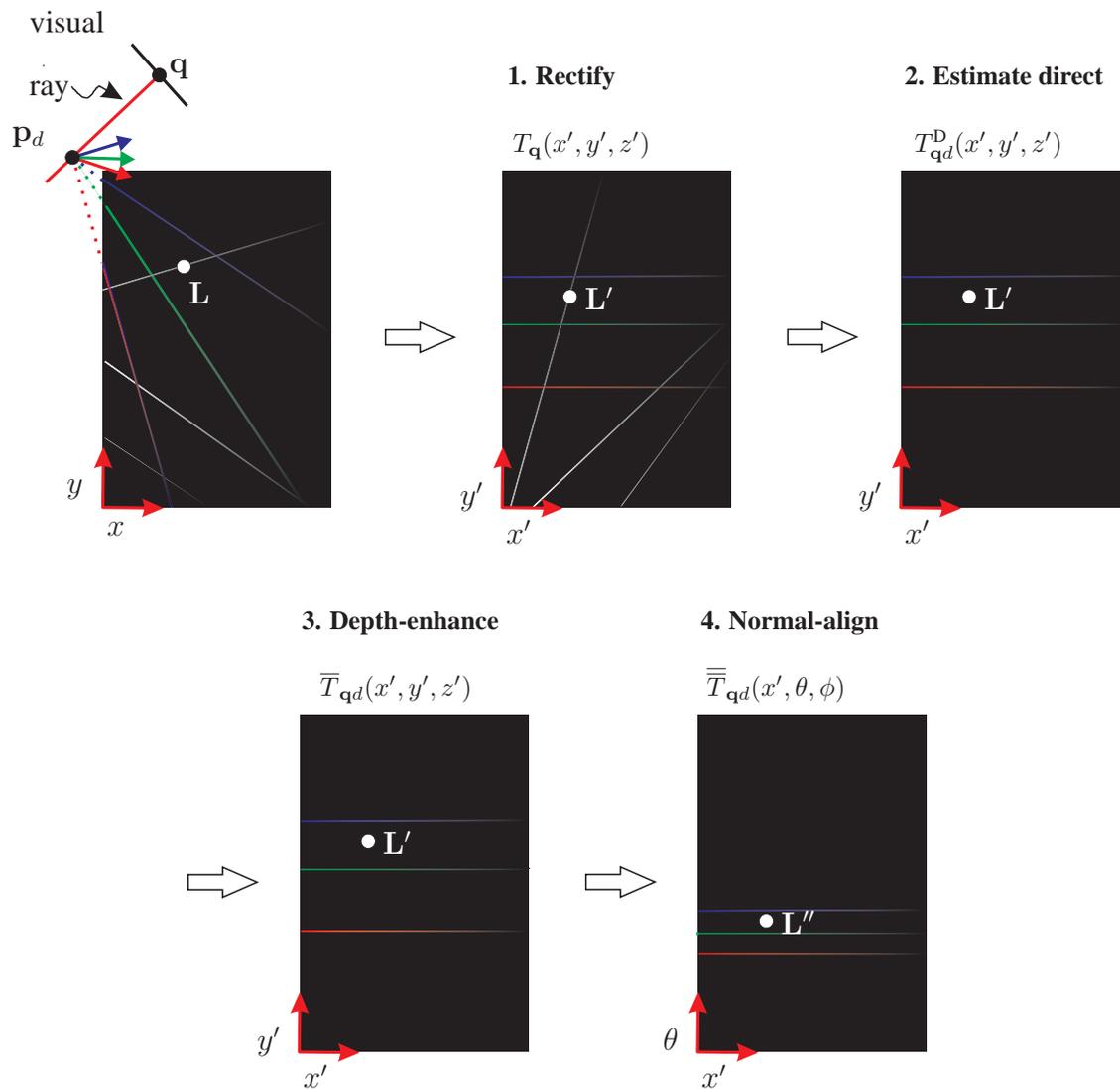
Figure 4.5: Basic steps in scatter-trace analysis. Each step corresponds to a transformation of the pixel's original scatter trace, for a specific depth hypothesis. Red, green and blue colors indicate three impulse scatter traces that contribute to the pixel $\mathbf{q}$'s direct scatter trace for the "true" depth d. The indirect scatter trace component is shown in white. Colored vectors in the leftmost figure are along the bisector of the viewing ray and the similarly-colored impulse scatter trace(see Section 4.4 for details and Figure 4.16, bottom right for a real-scene example).

- **Monotonicity:** Their intensity must decrease monotonically in a radial direction away from the convergence point.

Given a hypothesized depth for a pixel $\mathbf{q}$, it is possible to use these constraints to decompose its scatter trace into two components—an estimated direct component, $T_{\mathbf{q}d}^{\mathrm{D}}$, that is fully-consistent with both the hypothesized depth d and the above constraints, and a component that is not:

$$T_{\mathbf{q}} = T_{\mathbf{q}d}^{\mathrm{D}} + T_{\mathbf{q}d}^{\mathrm{I}} \ . \tag{4.4}$$

This observation, which forms the key idea of our shape recovery approach, allows us to assign a "consistency measure" to each depth hypothesis. In the single-view case, we use a measure that evaluates the consistency of the estimated direct component with pixel's entire scatter trace. When multiple views are available, our measure evaluates the mutual consistency of the estimated direct component at corresponding pixels in the input views. Since both measures depend on the problem of estimating the direct scatter trace, we consider this problem first.

4.3.1   Estimating the Direct Scatter Trace

In the absence of additional information about the scene, the decomposition of Eq. (4.4) is not unique.[2] In light of this ambiguity, we compute the most conservative estimate of the direct component, i.e., an estimate that is guaranteed to be at least as large as that of the "true" direct scatter trace when the hypothesis corresponds to the correct depth d^*:

$$T_{\mathbf{q}d^*}^{\mathrm{D}}(\mathbf{L}) \geq T_{\mathbf{q}}^{\mathrm{D}}(\mathbf{L}) \quad \text{for all } \mathbf{L} \ . \tag{4.5}$$

We do this by enforcing the convergence and monotonicity constraints in succession.

---

[2]For example, an estimated direct component that is zero everywhere trivially satisfies the three constraints, for any depth.

In particular, each depth d defines a unique 3D convergence point, $\mathbf{p}_d$, for all "streaks" of the direct component. These streaks must therefore lie on the pencil of rays through $\mathbf{p}_d$. To enforce the convergence constraint, we rectify the pixel's scatter trace so that all these rays become parallel to the x axis of the coordinate system. This is a linear projective warp that maps point $\mathbf{p}_d$ to the point at infinity along x. In two dimensions, it corresponds to standard epipolar image rectification [43] with the epipole at $\mathbf{p}_d$ (Figure 4.5-1).

Every line parallel to the x axis in this rectified scatter trace corresponds to a distinct ray through $\mathbf{p}_d$. Geometrically, the rectification operation converts the scatter trace's original spatial (x, y, z) coordinates into coordinates (x', y', z') that encode ray direction (coordinates y' and z') and position along the ray (coordinate x'). In this coordinate system, the monotonicity constraint tells us that the intensity of the direct component must be non-increasing along the x'-axis. Since we are dealing with discrete measurements, we enforce the constraint recursively by simply computing a "running minimum" across points of the rectified scatter trace in the x'-direction (Figure 4.5-2):

$$T_{\mathbf{q}d}^{\mathrm{D}}(\,\mathbf{L}') = \min\left(\,T_{\mathbf{q}}(\,\mathbf{L}')\,,\;T_{\mathbf{q}d}^{\mathrm{D}}(\,\mathbf{L}' - \mathbf{X}'\,)\,\right)\;,\tag{4.6}$$

where $\mathbf{L}'$ denotes a scatter trace point in rectified coordinates and $\mathbf{X}'$ is the unit vector along the x'-axis.

Note that the computation in Eq. (4.6) leaves unaffected any signal that decreases monotonically along the x'-axis. As such, it will not attenuate the pixel's "true" direct scatter trace when rectification occurs at the correct depth. Intuitively, this maximally-conservative, depth-based estimate of the direct scatter trace can be thought of as a scatter trace whose direct component is contrast-enhanced at the correct depth. This is because the "true" indirect scatter trace will typically not satisfy all three constraints (ray intersection, convergence, monotonicity) and, hence, will be attenuated for every depth hypothesis.

4.3.2   Single-View Shape from the Scatter Trace

If we know in advance that a pixel's indirect scatter trace is negligible (e.g., we know that the point projecting to the pixel is opaque and receives few inter-reflections), the above estimation procedure leads to a very simple single-view depth estimation algorithm—we just search for the depth along each pixel's viewing ray that best explains the scatter trace as a pure direct component.

In practice, we discretize the depths and evaluate the consistency between $T_\mathbf{q}$ and $T^\mathrm{D}_{\mathbf{q}d}$ for each pixel $\mathbf{q}$ and depth d. To do this, we first measure the point-wise consistency between these two scatter traces under the assumption of additive Gaussian noise with standard deviation σ,

$$W(d,\ \mathbf{L}') = \exp -\frac{1}{\sigma^2}\left[\, T_\mathbf{q}(\mathbf{L}') - T^\mathrm{D}_{\mathbf{q}d}(\mathbf{L}')\,\right]^2 , \tag{4.7}$$

and then use this consistency to enhance the direct component in the original scatter trace measurements (Figure 4.5-3):

$$\overline{T}_{\mathbf{q}d}(\mathbf{L}') \;=\; W(d,\ \mathbf{L}')\, T_\mathbf{q}(\mathbf{L}') . \tag{4.8}$$

In effect, this enhancement operation weighs each original scatter trace measurement by an upper-bound estimate of the likelihood that it was due to direct reflection. Our final metric aggregates the weighted measurements in Eq. (4.8) across all positions:

$$\mathcal{C}^1(d) \;=\; \sum_{\mathbf{L}'} \overline{T}_{\mathbf{q}d}(\mathbf{L}') . \tag{4.9}$$

The metric computes the total mass of measurements explained by a given depth d, under the condition that they are due to direct reflection. To assign depth, we maximize it.

4.4 Scatter-Trace Stereo

The indirect component of the scatter trace cannot be ignored when reconstructing scenes with transparent surfaces. This means that we cannot use the criterion in Eq. (4.9) to assess the validity of a given depth hypothesis. We therefore generalize our single-view analysis by evaluating the mutual consistency of scatter traces at corresponding pixels in two or more views.

In particular, let $\mathbf{q}_1, \mathbf{q}_2$ be a hypothesized correspondence between two pixels in a pair of views, let d be the depth implied by this correspondence, and let $\overline{T}_{\mathbf{q}_1 d}$ and $\overline{T}_{\mathbf{q}_2 d}$ be their rectified and depth-enhanced scatter traces (Eqs. (4.6)-(4.8)). To compare these two scatter traces, we first warp them in a way that makes point-wise comparisons meaningful and then simply compute their cross-correlation:

$$\mathcal{C}^2(d) \;=\; \sum_{\mathbf{L''}} \overline{\overline{T}}_{\mathbf{q}_1 d}(\mathbf{L''})\, \overline{\overline{T}}_{\mathbf{q}_2 d}(\mathbf{L''}) \,, \tag{4.10}$$

where $\mathbf{L''}$ denotes positions in warped coordinates and $\overline{\overline{T}}_{\mathbf{q}_1 d}, \overline{\overline{T}}_{\mathbf{q}_2 d}$ denote the "aligned" versions of $\overline{T}_{\mathbf{q}_1 d}, \overline{T}_{\mathbf{q}_2 d}$. At a superficial level, the metric in Eq. (4.10) can be thought of as a direct extension of traditional correlation-based stereo matching to multi-view scatter-trace photography.

The two outstanding questions are how to define the alignment warps in the two views, and under what conditions is the cross-correlation in Eq. (4.10) physically meaningful? We answer both questions by observing that the depth-enhanced scatter traces strongly constrain the normal(s) of the surface point $\mathbf{p}_d$, projecting to the two pixels.

Every line along the x'-axis of the rectified coordinate system corresponds to a unique ray of incidence at point $\mathbf{p}_d$. Given a pixel $\mathbf{q}_i$ and such a ray, there is a unique surface normal that specularly reflects light to pixel $\mathbf{q}_i$ from light sources on that ray (Figure 4.5, far left). This

normal is the bisector of the ray of incidence and the visual ray through pixel $\mathbf{q}_i$, and can be represented by two angles, $\theta_i(y', z')$ and $\phi_i(y', z')$.

A high value of the direct scatter trace for a position (x', y', z') along a specific ray of incidence can be interpreted as a "vote" for its corresponding normal. More generally, the depth-enhanced scatter trace along view i can be thought of as voting for the micro-facet distribution at point $\mathbf{p}_d$. To compare two such scatter traces point by point, we align them so that points in two different scatter traces correspond to the same normal. This alignment operation transforms rectified coordinates (x', y', z') into a surface-centered, $x'\theta\phi$-coordinate system that is defined in terms of the surface normal at the hypothesized depth (Figure 4.5-4):

$$\overline{\overline{T}}_{\mathbf{q}_id}(x',\ \theta_i(y', z'),\ \phi_i(y', z')) \;=\; \overline{T}_{\mathbf{q}_id}(x', y', z') \,. \tag{4.11}$$

In this coordinate system, the cross-correlation of Eq. (4.10) can be thought of as measuring the intersection of two micro-facet distributions. It will be maximized at the depth where there is maximal overlap between them, i.e., where their intersection maximally accounts for the estimated direct scatter traces in both views. The distribution itself can be computed by aggregating the votes for each normal at the depth d that maximizes Eq. (4.10), i.e.,

$$D(\theta, \phi) \;=\; \sum_{x'} \overline{\overline{T}}_{\mathbf{q}_1d}(x', \theta, \phi)\, \overline{\overline{T}}_{\mathbf{q}_2d}(x', \theta, \phi) \,. \tag{4.12}$$

Therefore, maximizing Eq. (4.10) leads to a highly-detailed description of the local surface projecting to each pixel—both a depth d and its micro-facet distribution.

4.5   Implementation Details

In this section we discuss how we practically implemented the scatter trace stereo algorithm and applied it to multi-camera reconstruction. We also discuss how to obtain sub-pixel depths

and then present a phase unwrapping technique to reduce the number of required images for a scatter trace.

4.5.1 Multi-view Reconstruction

We applied the stereo metric to a multi-view camera arrangement in order to further constrain it and reduce ambiguities especially in the presence of strong caustics.

For a given reference view i we compute the stereo metric between this view and all other views j, which is the cross-correlation of $\overline{\overline{T}}_{\mathbf{q}_i d}(x', \theta, \phi)$ and $\overline{\overline{T}}_{\mathbf{q}_j d}(x', \theta, \phi)$ for all pixels $\mathbf{q}$. This effectively gives us a metric response for a set of stereo volumes each of which has view i as the reference. The rectifications between each stereo pair will be different so we need to re-sample each stereo volume so they lie on the same discrete image coordinate space as the reference view i. Once we do this, we have metric responses along rays from each stereo volume. We can then use the cross-correlation between metrics as our multi-view metric. Figure 4.6 illustrates the improvement of a trinocular reconstruction over the stereo reconstruction from the same reference view. Notice how much smoother the depth map is. The recovered area in the trinocular version is also slightly expanded since an additional view is considered.

4.5.2 Sub-pixel Reconstruction

In addition, we perform reconstructions at sub-pixel resolution by upsampling *scatter trace slices*. A scatter trace slice $T_{\mathbf{Q}}$ is a 2D array of scatter trace responses for a row of pixels $\mathbf{Q}$ in the image as the light source moves to a 1D set of positions, see Figure 4.7. In a stereo system we have two such scatter trace slices, and so by increasing the resolution of the pixel row dimension we also increase the depth resolution. Scatter trace slices exhibit strong spatial

Stereo depth map   *Trinocular depth map*

Figure 4.6: Juice reconstruction results. This figure also illustrates the improvement in reconstruction between two and three view reconstructions using the same reference view.

coherence, since streaks correspond to traces of the reflection off of continuous surfaces. This allows us to use a form of anisotropic diffusion [65] in order to get a smooth interpolation that preserves image gradients. Figure 4.7 also shows the up-sampled scatter trace slice after anisotropic diffusion.

4.5.3   Efficient Capture by Phase Unwrapping

In order to make data capture efficient as well as to reduce storage space, we present another method for reducing the number of required images to obtain effective scatter traces. This reduction makes full object scanning far more practical.

So, rather than scanning an object by displaying a single moving light source, we illuminate it with ρ sources simultaneously (we use $\rho = 20$ in our experiments), moving each the small
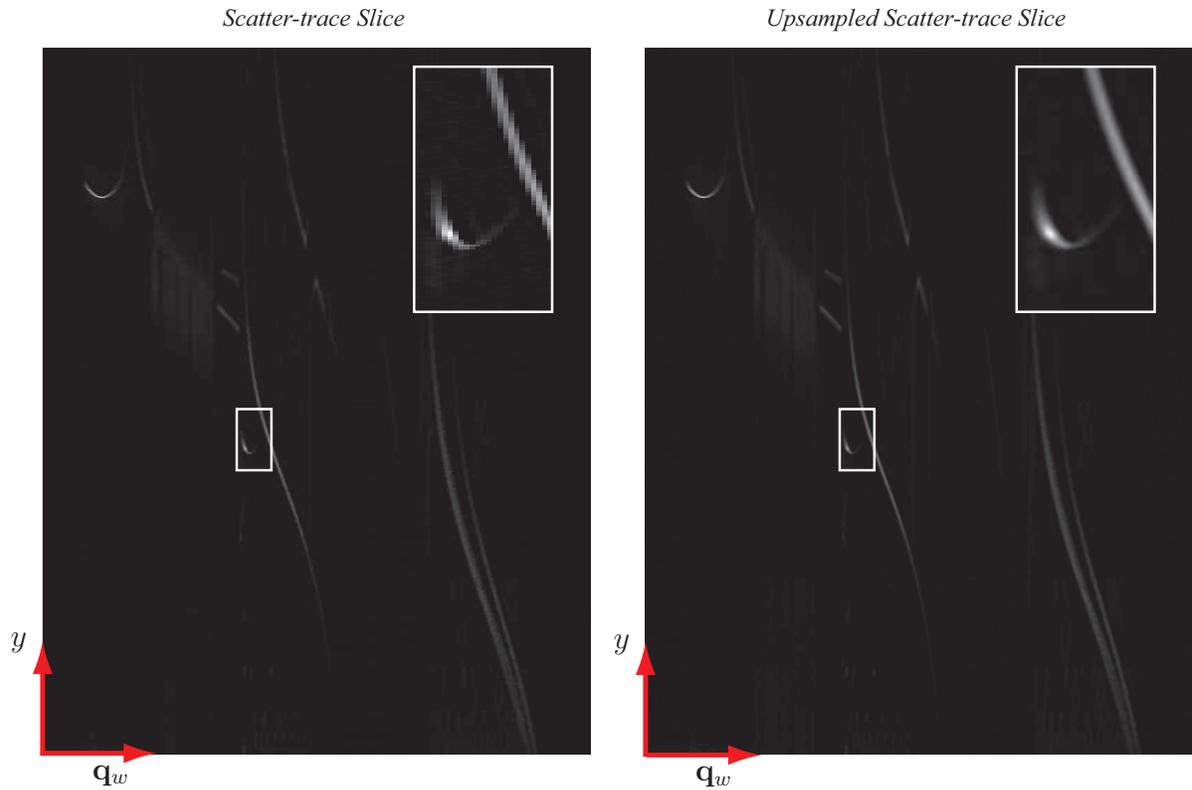
Figure 4.7: A scatter trace slice shows the response for each pixel on a row $\mathbf{q}_w$ and each light source position y. Left: Original slice. Right: Up-sampled slice by 4 in the pixel dimension w.

amount necessary to cover the same volume as the original scan. The captured trace we call $T_\mathbf{Q}^\mathrm{m}$. This results in far fewer images necessary to capture the same number of light source positions, however there are significant ambiguities generated in the scatter trace. This new scatter trace essentially consists of ρ copies of the original signal repeated across it. While the application of the scatter trace constraints (Section 4.3) through scatter trace stereo do constrain the metric to the part of $T_\mathbf{Q}^\mathrm{m}$ corresponding to the object's direct reflection, strong caustic effects become a much bigger problem when duplicated across the scatter trace. In order to combat these ambiguities, we capture a second scan $T_\mathbf{q}^\epsilon$, where the scatter trace volume is illuminated

by a larger light source that incrementally illuminates each region covered by the ρ previously displayed sources. This allows us to unwrap the phase of the first scan. So, our final scatter trace is composed as: $T_{\mathbf{q}}^{\tilde{}} = T_{\mathbf{q}}^{\mathrm{m}} \otimes T_{\mathbf{q}}^{\epsilon}$, where $\otimes$ represents element-wise multiplication (See Figure 4.8).
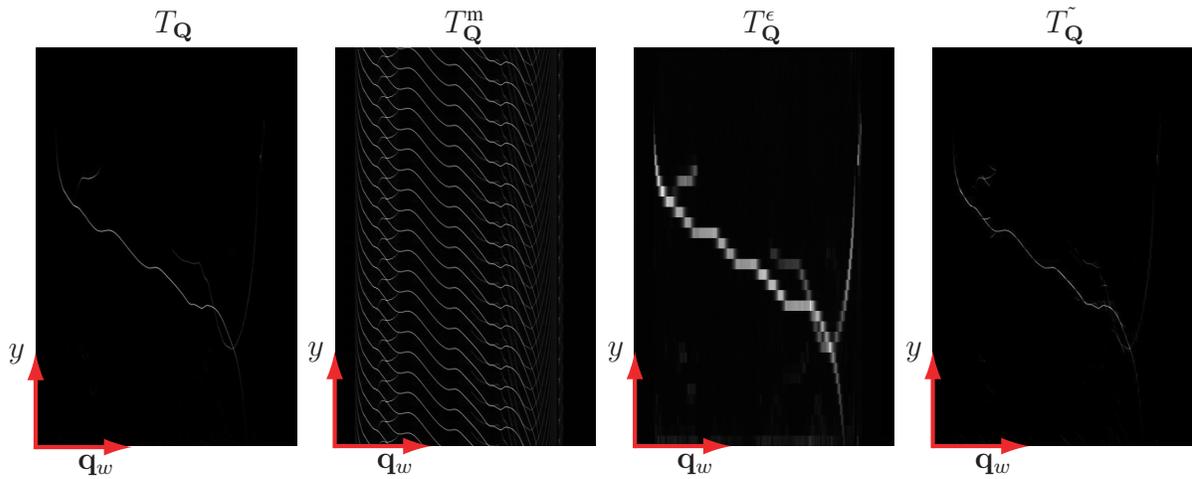
Figure 4.8: From left to right: The original full resolution scatter trace slice. Scatter trace slice with multiple sources. Unwrapping scatter trace slice. The reconstructed scatter trace resulting from $T_{\mathbf{q}}^{\mathrm{m}} \otimes T_{\mathbf{q}}^{\epsilon}$.

This method is significantly different from many coded pattern techniques, since these generally require threshold decisions to determine whether or not a pixel receives light from a particular source. We wish to avoid this type of decision, since the dynamic range of scatter traces may be very large, depending on the reflectance properties and internal light, so making a binary decision is not appropriate.

4.6   Experimental Results

We acquired 2D scatter traces using the acquisition setup in Figure 4.9, bottom right, with the camera about $1m$ away from the objects. A 2-pixel-wide vertical stripe on the LCD monitor acted as our light source, giving us 797 distinct positions that spanned a range of $41cm$ in the y-direction. The monitor was physically translated in the x-direction to 3-6 positions, spanning a range of 3-6cm depending on the scene. Instead of displaying vertical stripes individually, we used Schechner and Nayar's illumination multiplexing method [107]. This is also true of the efficient acquisition mode, where multiple light sources are each multiplexed in the same way. Objects were placed on a turntable and rotated by increments of $5°$ or $10°$ from their initial position to obtain additional views.
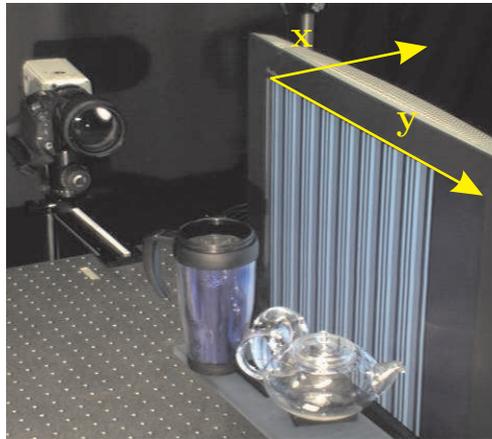
Figure 4.9: Acquisition setup

Since depth can be computed only for pixels that receive some light, pixels whose entire scatter trace was below an intensity threshold were pruned prior to reconstruction. We used a threshold equal to K times the median intensity across all scatter traces and all pixels. This "global" median can be thought of as providing an estimate of the noise level in the dataset—

since each pixel receives light from only a small subset of light source positions, intensities below this median are effectively due to noise. We used $K = 10$ for all experiments.

We assigned depth independently to each pixel in the first view by first evaluating metric $\mathcal{C}^2(d)$ in Eq. (4.10) at every possible integer disparity, and then "naively" choosing the disparity that maximized this metric. To assign a surface normal, we used the peak of the pixel's normal distribution, D, computed for that disparity.

Since we simultaneously recover both depth and normal information we look at the consistency between these shape measures to validate our results and as a natural outlier filter. Our approach was to locally fit planes to the geometry around each pixel to give geometry based normals. These normals were then directly compared to the normals we recovered from the scatter traces. We then reject surface points whose angular distance between the geometry based normals and the reconstructed normals was greater than $15°$.

We did not apply any other post-processing to the computed points and normals (e.g., smoothing, cleaning, etc). Our reconstruction procedure did not involve any tunable parameters.

Fish sculpture This scene represents an "easy" case for our method: its shape allowed a reasonably complete reconstruction because most pixels in its footprint received some illumination via direct specular reflection (Figures 4.1 and 4.10). Note that our reconstructions are based on a single stereo pair in this case and we used 1/8 pixel depth resolution. Despite the fact that all pixels were reconstructed independently, both the depth map and the normal map are highly uniform and capture fine surface details, including high-curvature regions near the beak and the eye.

Failures of the method correspond to (1) "missing" pixels, that were not reconstructed at all because they did not receive any light from the light source, and (2) reconstructed points

that deviated significantly from their "true" positions. The latter type of failure occurs at pixels near the object's silhouette, where the surface is viewed very obliquely, and at pixels where the direct reflection component has small magnitude. These pixels are readily identifiable because the magnitude of the matching criterion at the optimal depth, $\max_d \mathcal{C}^2(d)$, is very low (Figure 4.10). Although we did not attempt to do so, it should be possible to use this magnitude as a confidence measure for outlier rejection.

We compared our reconstruction with laser scans completed after coating the object with a white spray-on powder. Figure 4.10 shows the depth and normal maps side-by-side for these two methods as well as the difference images and a cross section of the reconstructions. The depth maps closely align, with an RMS error of 0.51mm and standard deviation of 0.35mm. The depths match well over large areas, however there are noticeable 'stepping' artifacts that correspond to areas of higher curvature and those regions where the scatter trace metric $\mathcal{C}^2(d)$ has a broad peak. The reconstructed normals match well in terms of low frequencies, however there is higher frequency noise in the laser scan that does not exist in our reconstructed normals (or on the true surface). We also computed the Poisson reconstruction that combines both the depth and normal data into a mesh [63] using default settings. The comparison with the laser depths show even closer alignment and most of the surface artifacts are gone. In addition the fine detailed features such as the ridges on the bottom of the fish are clearly recovered (shown in the insets in Figure 4.10 and meshes shown at the top of Figure 4.11).

Decorative bottle This scene had a curved transparent exterior with a painted interior with similar challenges to the *fish sculpture*. We present this dataset to show that the efficient capture process described in Section 4.5 works, and to demonstrate the metric quality of our reconstructions for a full model. The bottle has four distinct sides and we were able to recover trinocular depth maps of each side using scatter traces captured with the efficient method shown in Figure

4.8. We estimated depths using upsampled scatter trace slices giving us 1/4 pixel depth resolution. Figure 4.12 shows the depth and normal maps of one side of the bottle, as well as the reconstructed Poisson surface and comparisons to laser scan data. We aligned our recovered point set of the full object with the laser point set using a least squares best fit. Figure 4.13 show the 3D surfels for all four sides of the bottle combined as well as cross section comparisons of one side with the laser scan. Some parts of the top of the bottle had no visible direct reflections and only "phantom" points were recovered in these areas. The very bottom also exhibits error in the normal map since double reflections occurred off of the stand the bottle rested on. Also, there is some bias in the depth differences toward the top of the bottle, this may be due to misalignment between the reconstructions or errors in either the laser scan or our depths. Despite these issues, the four sides match well, showing that our method is able to recover, metric results without skew.

Juice bottle The primary challenges in this scene were the rather complex geometry of the bottle's upper section and the presence of an opaque label on the surface. Our method was able to reconstruct the label's surface quite well, albeit with more noise compared to the regions of exposed glass. While depths and normals in those regions were reconstructed well, the presence of self-occlusions and high curvatures meant that relatively few visible surface points reflected light toward the camera. This caused significant gaps in the reconstruction, which would require additional viewpoints to complete. In this respect, our approach is significantly less efficient than techniques for reconstructing opaque non-specular scenes, where almost all surface points visible to a pair of views can be reconstructed in one step. The depth and normal maps presented show our reconstructions using trinocular stereo and the surfels are the combination of three such trinocular reconstructions (Figure 4.14).

Multiple objects This was by far the most challenging scene, with a variety of complex light

transport phenomena and occlusions. Three observations can be made about our results. First, despite this complexity, we obtained detailed reconstructions for significant parts of the scene, including small high-curvature structures (e.g., a portion of the teapot lid).

Second, the specular reflectance properties of the scene's opaque regions were not sufficient to enable their reconstruction in this case, because of the low magnitude of their scatter trace. Third, the high curvature of the interior bubble and the lack of direct reflection at pixels in its footprint led to the reconstruction of a "phantom" surface. This artifact can be mitigated by collecting additional views that cause more points on the exterior surface to reflect light directly toward the camera. More generally, however, we believe that correct treatment of interior structures should involve reasoning analogous to occlusion handling in multi-view stereo, where depth hypotheses are analyzed globally, and reconstructed front-to-back (or outside-in).

The reconstructions in Figure 4.15 are obtained from a stereo pair of views and depth estimated at pixel resolution.

4.7   Concluding Remarks

A key contribution of our work is to show that, despite the highly-complex optical properties of inhomogeneous transparent scenes, accurate reconstruction is indeed possible with simple algorithms. Looking forward, we believe that even more general reconstruction problems are now coming within reach. We are currently investigating the reconstruction of complete surface models in the presence of occlusions, and the reconstruction of scene interiors, along with their surface.
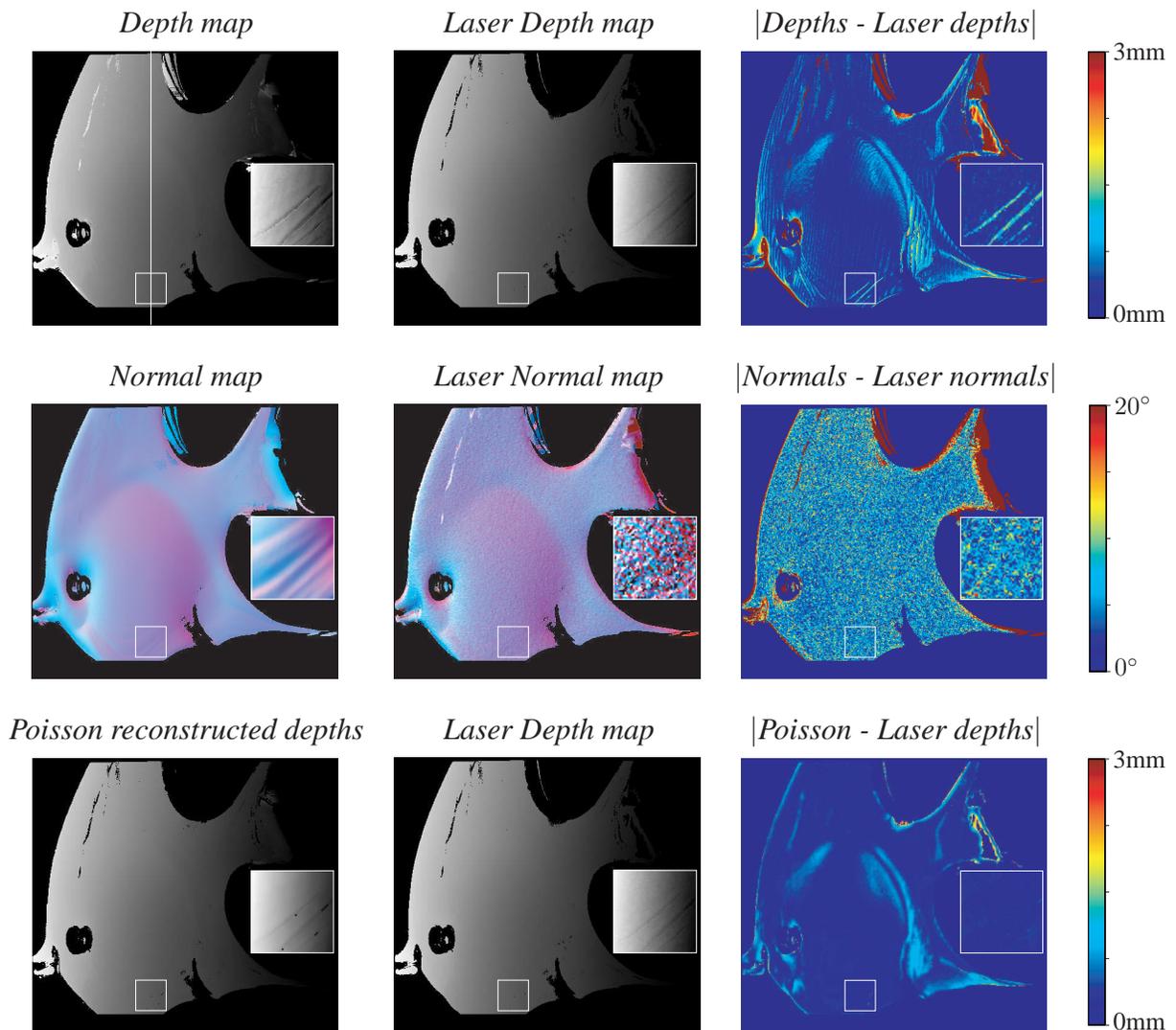
Figure 4.10: Fish reconstruction results using stereo and 1/8 pixel depth resolution. The results show a striking improvement over the basic depth map when the depths and normal are combined in the Poisson reconstruction. Note the detail in the inset region, where the ridge detail is exaggerated and noisy in the depth map but the Poisson map is smoothed by the accurate normals and closely matches the laser depths. The insets show how our recovered normals are much cleaner than the laser normals and are able to improve the depth map when combined.

Depth Reconstruction   *Laser Reconstruction*   *Poisson Reconstruction*

$\max_d \mathcal{C}^2(d)$ *3D Surfel view*

Cross sections of Depth map, Poisson reconstruction and Laser depth map

Fish top

Depth map
Laser depths
Poisson Reconstruction
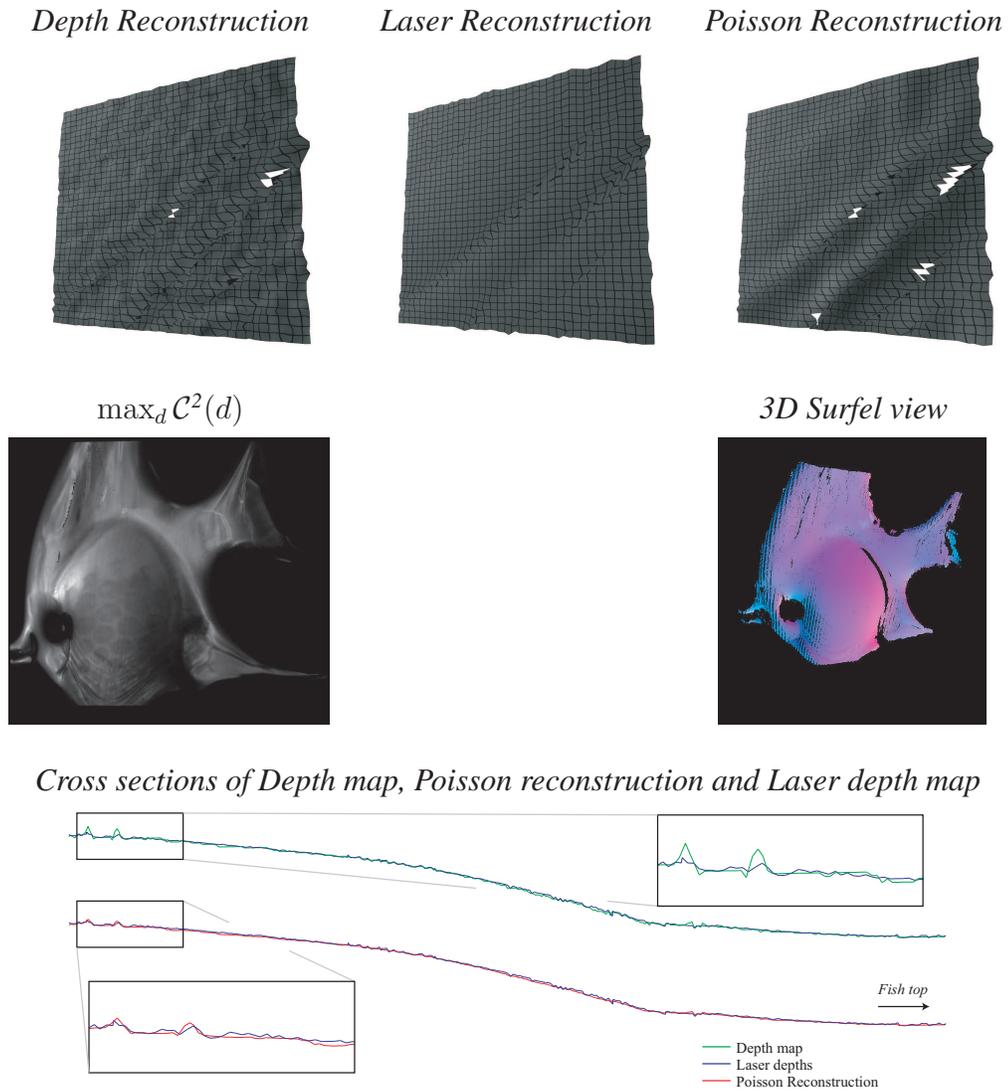
Figure 4.11: Top: Mesh reconstructions showing the inset regions from Figure 4.10. Notice how the Poisson surface preserves more detail than the laser reconstruction with less noise. Cross sections are from column indicated on depth map (top left Figure 4.10), showing the depth map aligned with the laser depths and the Poisson map also compared with the laser depths. These show good alignment between the Poisson surface and the laser scan, with improved smoothness in the Poisson surface over the laser depths.
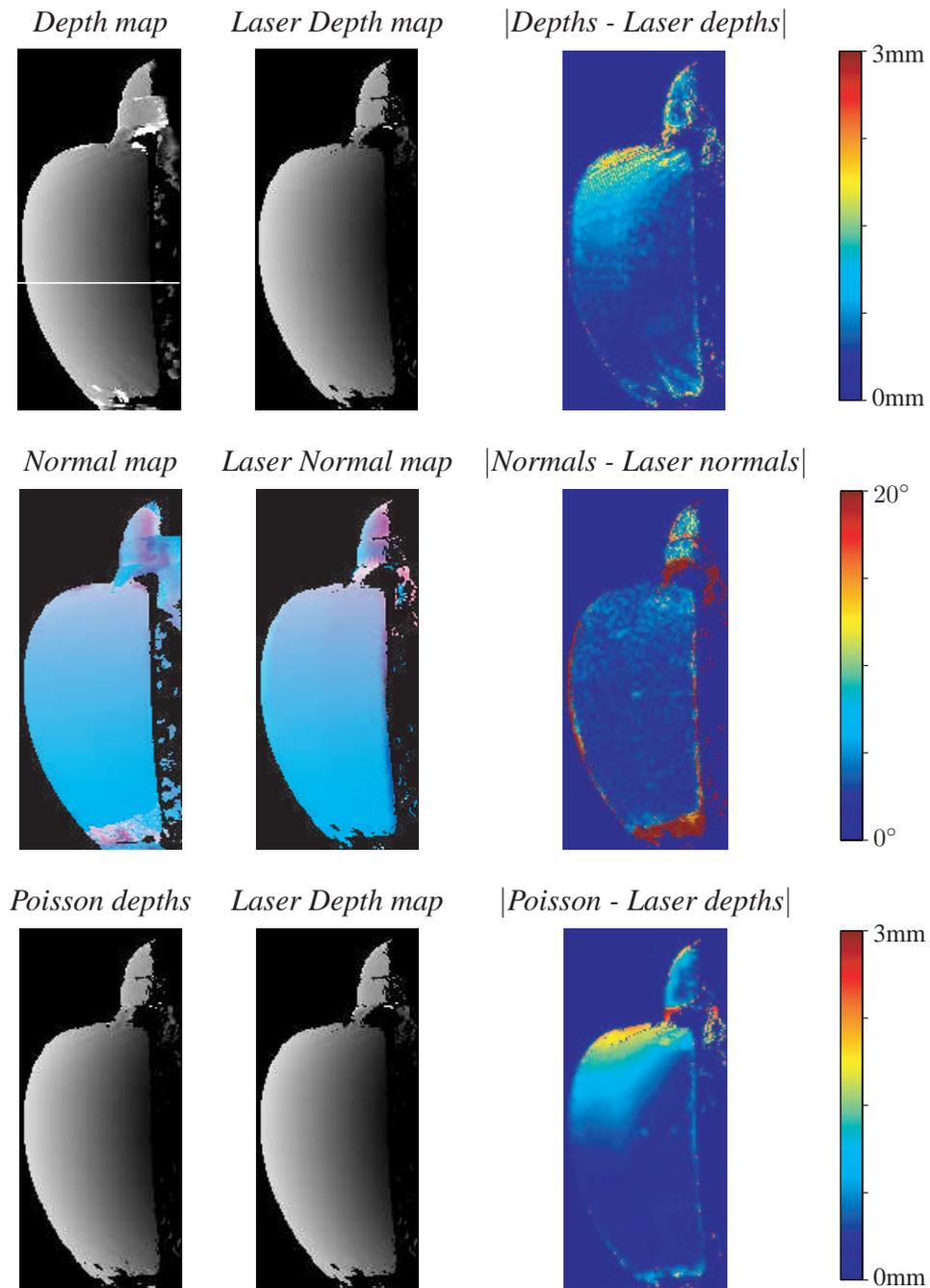
Figure 4.12: Reconstruction results for decorative bottle using high speed capture, trinocular stereo and 1/4 pixel depth resolution.

$$\max_d \mathcal{C}^2(d) \qquad \textit{3D Surfel view}$$

Cross section

Depth map
Laser depths
Poisson Reconstruction
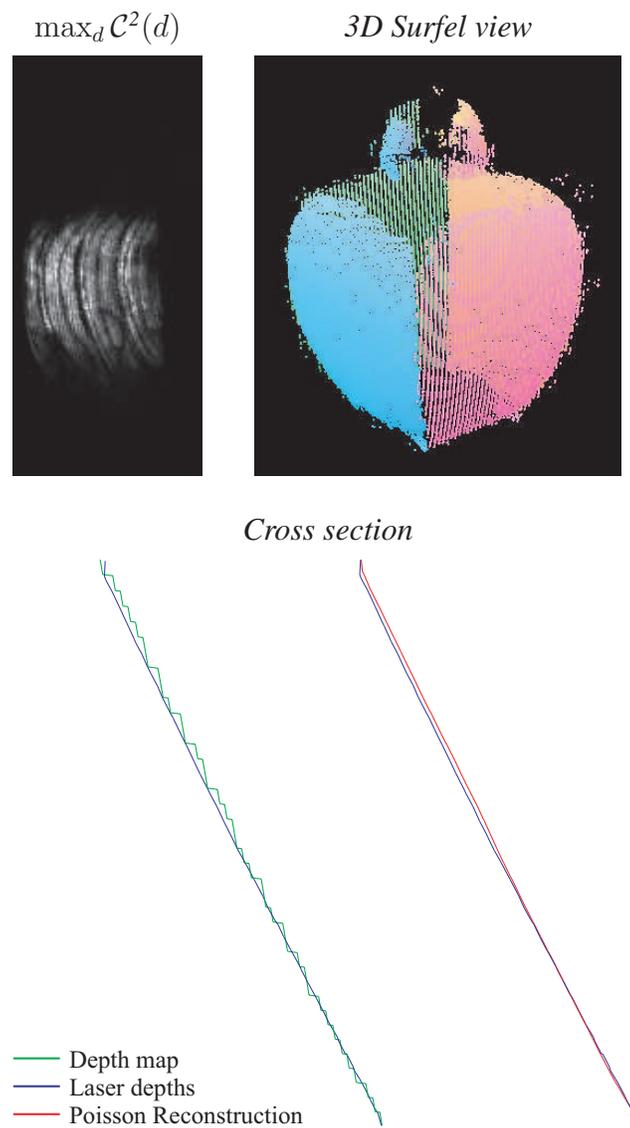
Figure 4.13: Reconstruction results for decorative bottle using high speed capture, trinocular stereo and 1/4 pixel depth resolution. 3D surfel view shows all four reconstructed sides together. The cross section is from row indicated on depth map (Figure 4.12 top left) and the left cross section shows a comparison between depths and laser depths and the right compares the Poisson surface with laser depths, showing improved alignment and smoothness.

Depth map $\max_d \mathcal{C}^2(d)$ *Normal map* *3D Surfel view*
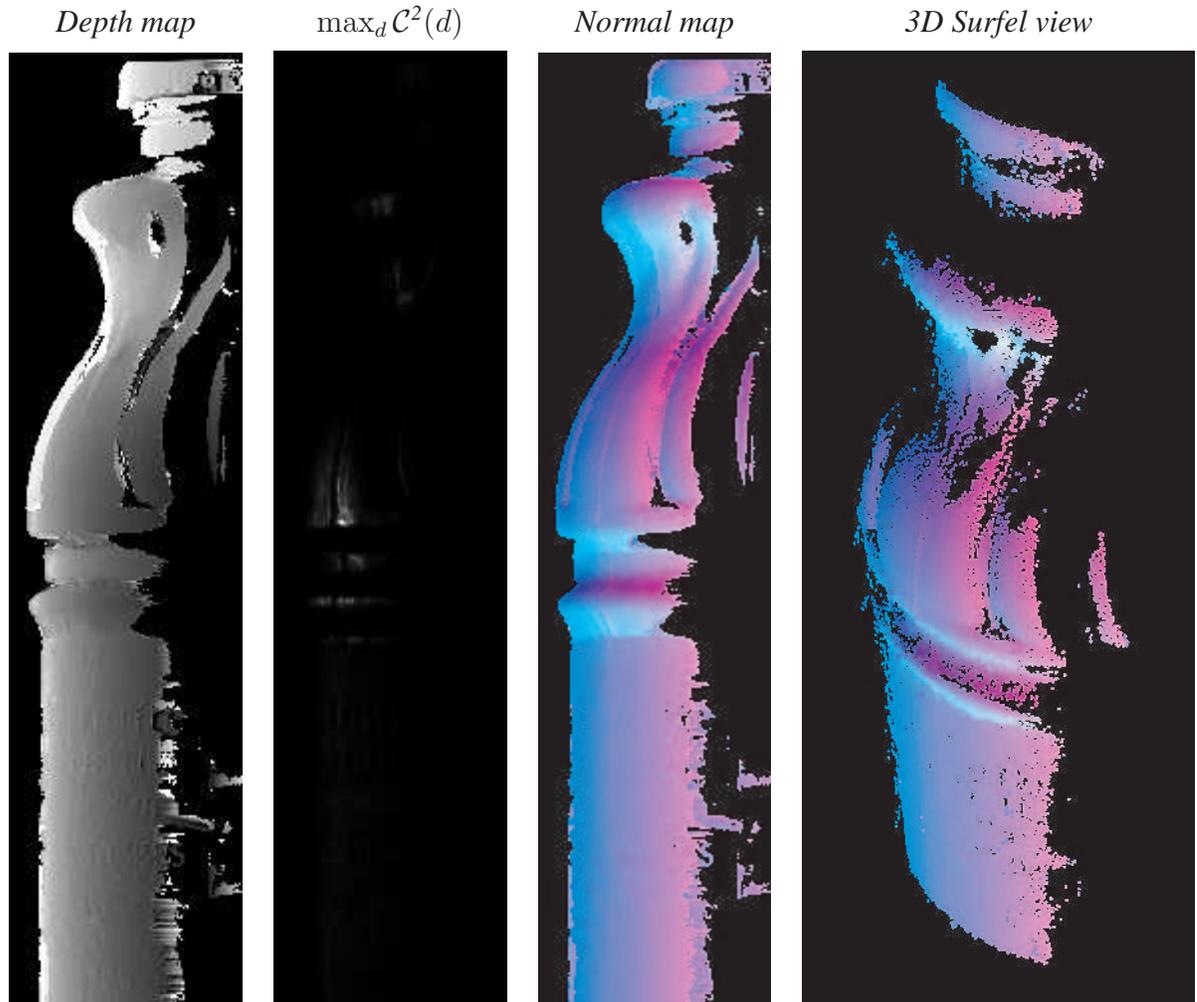
Figure 4.14: Reconstruction results for juice bottle using trinocular stereo and 1/4 pixel depth resolution. The 3D surfel view combines three neighboring trinocular datasets.

Depth map $\max_d \mathcal{C}^2(d)$

Normal map *3D Surfel view*
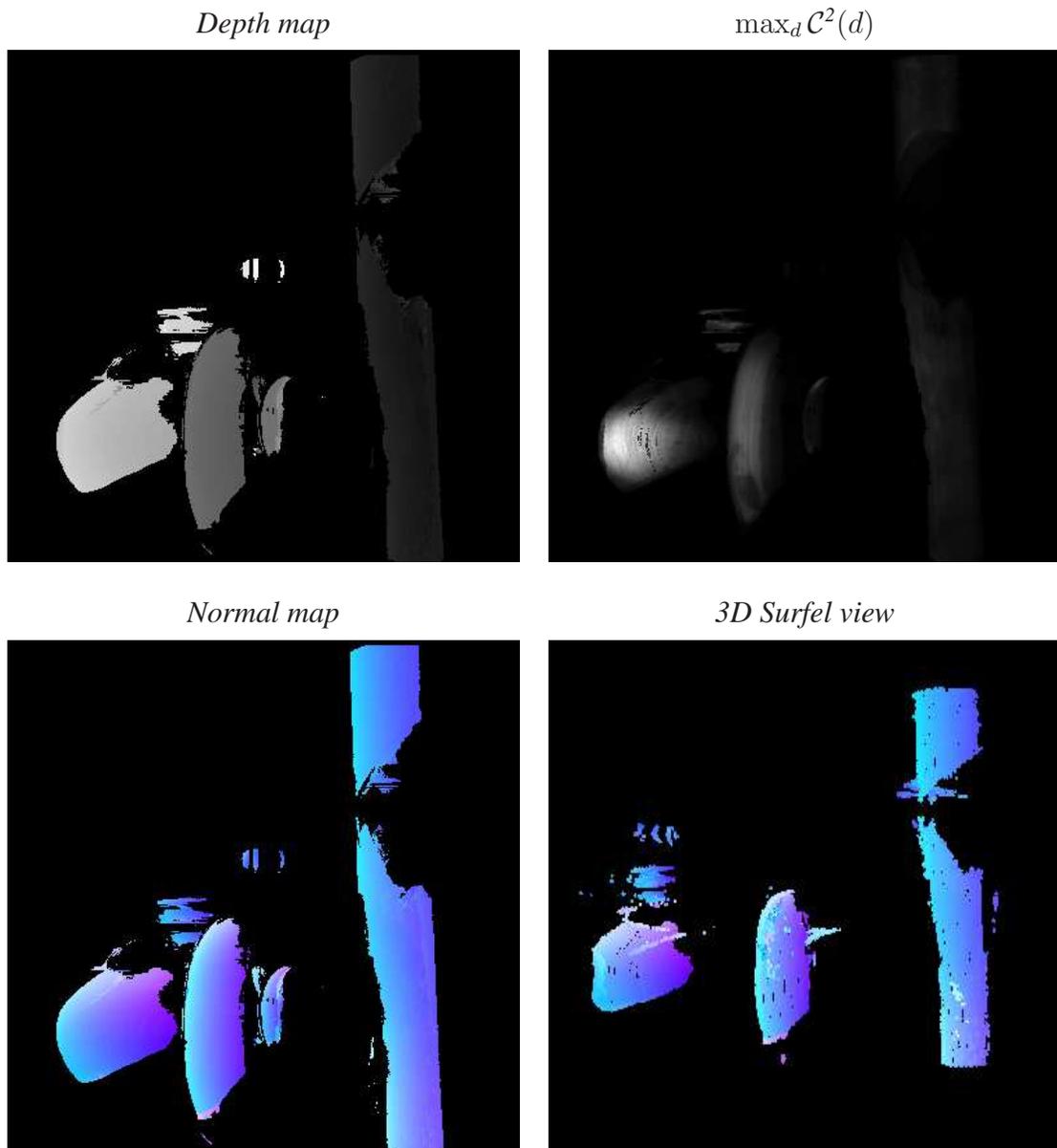
Figure 4.15: Reconstruction results for multiple object scene using a stereo view pair and pixel depth resolution.
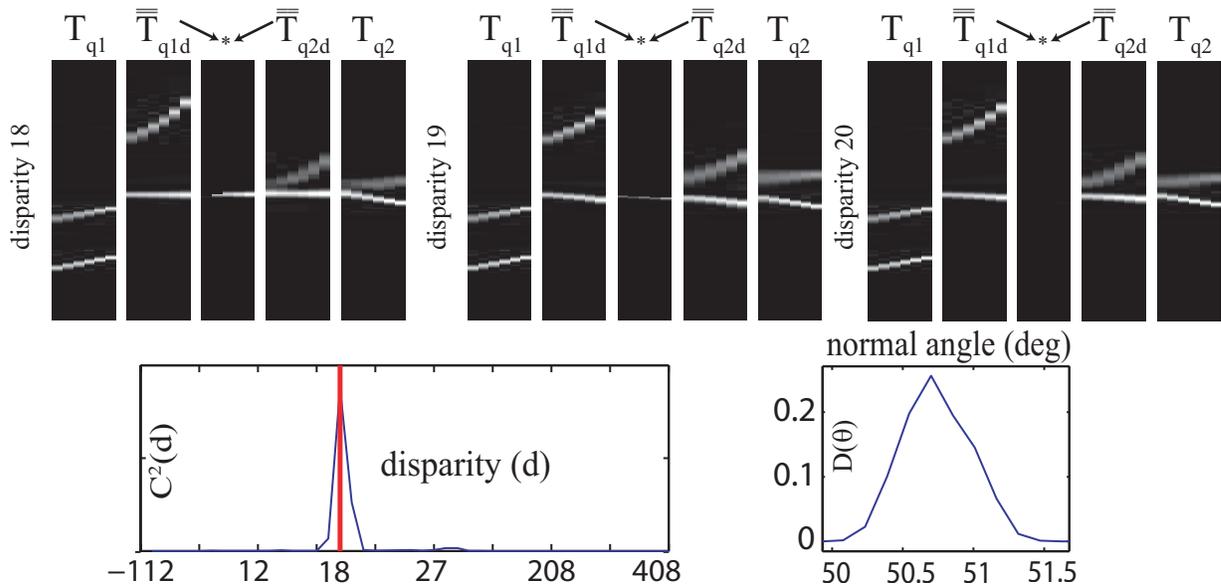
Figure 4.16: Scatter traces and intermediate results of applying the scatter trace stereo algorithm to the highlighted pixel in Figure 4.1(left). Scatter trace resolution was 6×797. Note the double streaks, corresponding to a direct reflection component and a secondary component due to indirect reflection (this occurs for almost all pixels in this scene). Also shown are the values of the matching criterion across disparities and the recovered normal distribution (only one angle is shown). Note the unambiguous peak at the true disparity, $d = 18$, despite the presence of strong secondary illumination.

Chapter 5

Depth from Reflectance Magnification

"Don't stare into a mirror when you are trying to solve a problem."

-Mason Cooley

5.1   Introduction

The problem of recovering the 3D shape of scenes that contain a wide variety of materials from diffuse to highly specular using simple and accessible equipment remains an open problem in computer vision. Previous work in this area can be divided into two classes, those that attempt to extend Lambertian approaches to non-Lambertian using stereo-based methods [19, 24, 39, 141] or photometric stereo or monocular methods [4, 5, 46, 48, 77] and those that build on techniques designed for recovering specular objects [18, 33, 85]. There is little reason to expect scenes to be exclusively Lambertian or specular and while progress has been made in both these areas, there is no one method that spans the full range from Lambertian to specular with a single, simple capture setup. Such scenes are especially challenging because

94

there is often a high dynamic range of brightness [26] and the appearance varies significantly depending on viewpoint and with changes in illumination. Thus it is difficult to make assumptions about the reflectance that holds for all the materials. Furthermore, the complexity of the capture process is a factor that varies significantly between 3D reconstruction techniques, with some requiring multiple views [24, 85, 141], exemplar materials [46], light source calibration [18, 85, 141], polarization filters [33, 77], hemispherical illumination gantries [77], and a significant quantity of images [18, 48, 85].

Motivated by these challenges, we present a novel approach that seeks to meet the following goals:

- **Invariance to the BRDF:** The approach should put as few restrictions on the type of material as possible and be able to recover diffuse as well as specular materials.

- **Simple capture setup:** The capture process should be straightforward, without specialized equipment and requiring only simple calibration.

- **Fast capture:** Only a few images should be required.

In this work we present two algorithms that use the same experimental setup shown in Figure 5.1, one requiring calibration and the other calibration free. The input to both algorithms is the same, a small number of images are captured from a single camera of the scene as it is illuminated by a near-field non-uniform planar light source (an LCD monitor) which displays a sequence of patterns. The source is then translated perpendicularly to its plane, thus altering its distance to the scene and a second set of images are captured. We define the perpendicular distance of a scene point from the near source plane as *source-relative depth*. Our uncalibrated approach is analogous to depth from defocus [98, 115] but rather than refocusing, we use the
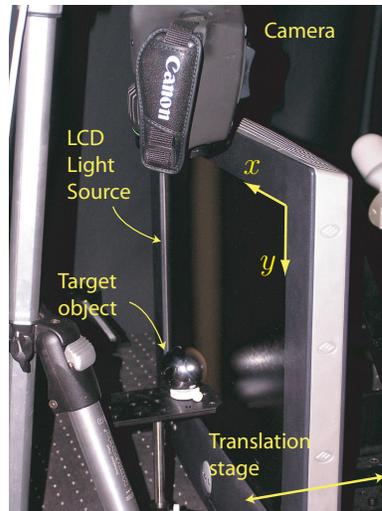
Figure 5.1: Experimental setup.

change in illumination between the translated source to compute the source-relative depth. In the calibrated algorithm, the orientation of the source to the camera is known and so for each pixel we recover a 3D point on the scene surface.

While we seek to be invariant to all BRDFs the range of reconstructable BRDFs differ between the calibrated an uncalibrated algorithms. In the uncalibrated version, we limit reconstruction to BRDFs that are smoothly varying and without strong skew in exchange for greater efficiency in capture. In contrast the calibrated version is able to reconstruct non-smooth BRDFs including specular objects. Both approaches are limited by the following. If the light source is partially occluded from the perspective of a point on the scene it is possible that that point will be incorrectly reconstructed. The light source size constrains the amount of the scene we are able to recover, since viable points must reflect light from the source back to the observer.

Our approach is related to BRDF invariant reconstruction methods that leverage distinctive properties of BRDFs [5, 24, 39, 46, 48, 141]. However many of these approaches rely on

point light source illumination for their invariants. For instance, the use of Helmholtz reciprocity [141] is not practical for specular materials because the reciprocity only applies to pairs of camera and point light source positions. In this case only specular surface points with a particular normal can be recovered. The same problem occurs when image ratios are used as in [24] because the assumption is based on changing the illumination of a point source and enough light must be reflected toward all cameras to measure the ratio. The approach of using exemplar materials [46] also breaks down when specular objects are considered, since the only change in appearance would be at the specular highlights when the light source position is changed.

Dense arrays of point light sources have been used to overcome the above problems in [4, 5, 18, 33, 48], however these methods are based on photometric stereo and either make assumptions about the symmetry of the material's BRDF [4, 5, 48] or only apply to glossy or specular materials [18, 33]. The output from these algorithms is a normal map rather than a depth map or point cloud as in our approach. In this respect our work is closer to those algorithms that are based on specular stereo. These algorithms work by observing specular highlights on the target surface, allowing depth and normals to be recovered by triangulation [13, 15, 102, 133, 137]. As noted by [102], this triangulation requires either knowledge of the of incident ray direction or a second view that can verify a hypothesized depth and normal. Rather than using multiple views, in our calibrated algorithm we use the translated light source to identify the incoming direction of rays in a similar way to [70, 85]. However we go beyond all these approaches to demonstrate that even highly diffuse materials along with specular materials can also be simultaneously reconstructed using a similar capture setup to the specular approaches.

Our work is closely related to [68] where a translated near-field point light source is used to acquire depth cues for Lambertian scenes. While we use a translated planar source, there

are similarities such as that both approaches are improved by increasing the distance the source is translated, and both techniques provide uncalibrated depth cues of the scene by observing the effect of reflectance changes as the light source is moved. While the method in [68] is not strictly normal independent our approach is, and while they do apply their method to non-Lambertian materials, this requires a highlight filtering stage rather than being invariant.

5.2 Reflectance Magnification

Consider a scene that is illuminated by a planar, spatially non-uniform light source (*e.g.*, an LCD monitor). We assume that the scene is viewed under perspective projection from a fixed viewpoint and that the scene is static, with an unknown 3D shape, and an unknown and spatially-varying BRDF.

With the scene static and the viewpoint fixed, there is a one-to-one correspondence between points in the image and points in the scene. Our goal is to recover two shape quantities independently for each image point:

- *Source-relative depth:* the distance of the corresponding scene point from the plane of the light source;

- *Camera-relative depth:* the distance of the image point from its corresponding scene point.

5.2.1 The 4D Reflectance Function for a Proximal Planar Source

The 4D reflectance function was introduced by Debevec *et al* [25] to represent scene appearance from a fixed viewpoint under directional lighting. In its original definition, this function
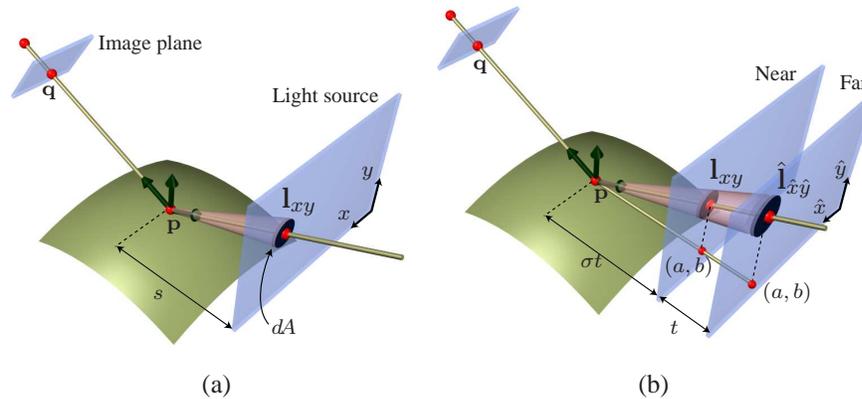
Figure 5.2: Imaging geometry. (a) Quantities involved in expressing the radiance incident at image point $\mathbf{q}$ due to a patch dA on the light source plane. (b) Our basic acquisition setup involves capturing images of the scene for two distinct positions—near and far—of a planar light source.

provides the radiance, $R_{\mathbf{q}}(\theta, \phi)$, arriving at an image point $\mathbf{q}$ when the scene is illuminated from direction (θ, ϕ) on the unit sphere. We begin by extending this definition to the case where illumination originates from a planar, spatially non-uniform source near the scene.

Let (x, y) be the coordinates of a point on the light source plane, expressed in the source's internal 2D coordinate system (Figure 5.2a). We model outgoing radiance at (x, y) as the product

$$I(x, y)\, L(\theta, \phi) \,, \tag{5.1}$$

where $I(x, y)$ is a unit-less modulation factor that describes variations in radiant exitance across the light source plane, and $L(\theta, \phi)$ is a spatially-independent radiance factor that represents the source's directional emission characteristics. In practical terms, the modulation factor is simply the pattern displayed on an LCD monitor; the radiance factor, on the other hand, describes the directional emittance characteristics of a single monitor pixel, under the assumption that all pixels emit light the same way.

The radiance received at an image point $\mathbf{q}$ due to light from a differential patch at position (x, y) on the light source is equal to

$$\lim_{\|A\| \to 0} \int_A \frac{f_{\mathbf{p}}(\mathbf{q} - \mathbf{p}, \ \mathbf{p} - \mathbf{l}_{xy}) \ L(\mathbf{p} - \mathbf{l}_{xy})}{\|\mathbf{l}_{xy} - \mathbf{p}\|^2} \ c(\mathbf{p} - \mathbf{l}_{xy}) \ I(x, y) \ dx dy \ , \qquad (5.2)$$

where $\mathbf{p}$ is the point projecting to $\mathbf{q}$; A is a patch centered at (x, y); $\mathbf{l}_{xy}$ is the 3D point corresponding to source position (x, y); $f_{\mathbf{p}}$ is the BRDF at $\mathbf{p}$; and $c()$ is a foreshortening factor that accounts for the normals of the source plane and the surface at $\mathbf{p}$.[1]

Intuitively, the integrand in Equation (5.2) describes the image contribution of light that leaves point (x, y) on the source, is received directly at scene point $\mathbf{p}$, and is then reflected toward point $\mathbf{q}$ on the image plane.

When the integrand is differentiable, the radiance in Eq. (5.2) reduces to

$$\boxed{\frac{f_{\mathbf{p}}(\mathbf{q} - \mathbf{p}, \ \mathbf{p} - \mathbf{l}_{xy}) \ L(\mathbf{p} - \mathbf{l}_{xy}) \ dA}{\|\mathbf{l}_{xy} - \mathbf{p}\|^2} \ c(\mathbf{p} - \mathbf{l}_{xy})} \ I(x, y) \ \overset{\text{def}}{=\!=} \qquad (5.3)$$

$$\boxed{R_{\mathbf{q}}(x, y)} \ I(x, y) \qquad (5.4)$$

where dA is the differential area element on the plane of the light source. In general, none of the quantities in Eq. (5.3) will be known for an unknown scene, except for the displayed pattern, $I(x, y)$.

The function $R_{\mathbf{q}}(x, y)$ generalizes the 4D reflectance function in [25] to the case of a 2D proximal light source. It is also four dimensional, with two dimensions ranging over the image plane and two ranging over the light source. For a fixed image point, $R_{\mathbf{q}}$ can be thought of as a 2D "reflectance image" that describes the point's appearance under illumination originating from every possible position on the source.

---

[1]Specifically, if ρ_1, ρ_2 are the angles that vector $\mathbf{p} - \mathbf{l}_{xy}$ forms with $\mathbf{p}$'s normal and with the normal of the light source plane, respectively, then $c(\mathbf{p} - \mathbf{l}_{xy}) = \cos \rho_1 \cos \rho_2$.

Below we consider how this image is affected by a translation of the light source plane away from the scene under the assumption of a differential BRDF (*i.e.*, where Eq. (5.3) applies). We revisit the case of a pure-specular BRDF, which is a delta function and thus non-differentiable, in Section 5.4.2.

5.2.2 Geometry of a Translating Planar Source

Suppose we move the planar light source away from the scene by translating it along its normal by a displacement t (Figure 5.2b). This translation will change the reflectance function of the scene, assigning a new, "far" reflectance image, $\hat{R}_\mathbf{q}$, to point $\mathbf{q}$.

The key observation in our approach is that the relation between the "near" and "far" reflectance images at $\mathbf{q}$ depends only the point's source-relative depth and is independent of the BRDF and the emission properties of the source. In particular, the far reflectance image can be thought of as a magnified and uniformly-dimmed version of the near reflectance image, with both magnification and dimming determined by the point's source-relative depth. We first express the source-relative depth as a fraction σ of the translation between the near and far planes:

$$s = \sigma t \, .$$
<div align="right">(5.5)</div>

Observation 1 (Reflectance Magnification) *If the light source has an infinite spatial extent, then for every point $(\hat{x}, \hat{y})$ on the far reflectance image, there is a point (x, y) on the near reflectance image such that*

$$\hat{R}_\mathbf{q}\left(\hat{x}, \hat{y}\right) = R_\mathbf{q}\left(x, y\right) \left(\frac{\sigma}{\sigma + 1}\right)^2 ,$$
<div align="right">(5.6)</div>

and

$$\begin{bmatrix} \hat{x} \\ \hat{y} \\ 1 \end{bmatrix} = \begin{bmatrix} (\sigma+1)/\sigma & 0 & -a/\sigma \\ 0 & (\sigma+1)/\sigma & -b/\sigma \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{5.7}$$

where (a, b) are the coordinates of the orthogonal projection of $\mathbf{p}$ onto the source plane, expressed in the source's coordinate system.

Proof Using Eq. (5.3) to expand Eq. (5.6) and re-arranging, we get:

$$\left(\frac{\sigma}{\sigma+1} \right)^2 = \frac{\hat{R}_{\mathbf{q}}(\hat{x}, \hat{y})}{R_{\mathbf{q}}(x, y)} = \frac{f_{\mathbf{p}}(\mathbf{q}-\mathbf{p},\ \mathbf{p}-\hat{\mathbf{l}}_{\hat{x}\hat{y}})\ L(\mathbf{p}-\hat{\mathbf{l}}_{\hat{x}\hat{y}})\ c(\mathbf{p}-\hat{\mathbf{l}}_{\hat{x}\hat{y}})}{f_{\mathbf{p}}(\mathbf{q}-\mathbf{p},\ \mathbf{p}-\mathbf{l}_{xy})\ L(\mathbf{p}-\mathbf{l}_{xy})\ c(\mathbf{p}-\mathbf{l}_{xy})} \cdot \frac{\|\mathbf{l}_{xy}-\mathbf{p}\|^2}{\|\hat{\mathbf{l}}_{\hat{x}\hat{y}}-\mathbf{p}\|^2}, \tag{5.8}$$

where $\hat{\mathbf{l}}_{\hat{x}\hat{y}}$ denotes the 3D point corresponding to position $(\hat{x}, \hat{y})$ on the far plane.

From similar triangles, points (x, y) and $(\hat{x}, \hat{y})$ on the near and far source planes, respectively, lie on the same 3D ray through point $\mathbf{p}$. It now follows that the BRDF, radiance, and foreshortening terms in Eq. (5.8) cancel, making the fraction dependent only on the ratio of squared distances from $\mathbf{p}$. Using similar triangles again, we get

$$\frac{\|\mathbf{l}_{xy}-\mathbf{p}\|^2}{\|\hat{\mathbf{l}}_{\hat{x}\hat{y}}-\mathbf{p}\|^2} = \left(\frac{\sigma}{\sigma+1} \right)^2 . \quad \blacksquare \tag{5.9}$$

Reflectance magnification tells us that the transformation relating the near and far reflectance images depends on just four parameters—three of which determine the 3D position of $\mathbf{p}$ uniquely (the source-relative depth and $\mathbf{p}$'s projection on the light source plane) and one that depends on the light source (light source displacement). This suggests a very simple BRDF-invariant algorithm for computing 3D shape: (1) capture the scene's reflectance function for a near and a far source position, (2) compute the three transformation parameters relating the near and far reflectance images of every point $\mathbf{q}$ on the image plane, and (3) convert these parameters into a 3D shape. Note that this algorithm does not rely on any camera calibration information—its only requirements are that the light source plane undergoes

translation along its normal; indirect illumination has a negligible effect on scene appearance (*e.g.*, inter-reflections, transparency and sub-surface scattering); and the source's spatial extent is sufficiently large.

Although it is possible to implement this algorithm directly, this would be very inefficient because it involves capturing the scene's full 4D reflectance function. For instance, an $N \times N$-pixel LCD acting as the light source would require capturing $2N^2$ images in total, *i.e.*, one image per pixel per source position. We therefore focus on ways to take advantage of the same basic principle but with far fewer images captured.

5.2.3 Speedup by Integration

One way to reduce acquisition requirements is to capture line integrals of the reflectance function rather than the function itself. This is possible because the dimming and magnification relations in Eqs. (5.6) and (5.7) are preserved under integration:

Observation 2 (Magnification of Reflectance Integrals) *If $R_{\mathbf{q}}^x(y)$ and $\hat{R}_{\mathbf{q}}^{\hat{x}}(\hat{y})$ are the axis-aligned integrals of $R_{\mathbf{q}}(x, y)$ and $\hat{R}_{\mathbf{q}}(\hat{x}, \hat{y})$ respectively, then*

$$\hat{R}_{\mathbf{q}}^x(\hat{y}) \;=\; R_{\mathbf{q}}^x(y) \left(\frac{\sigma}{\sigma+1} \right) . \tag{5.10}$$

See the Appendix for the derivation of Eq. (5.10). The above observation reduces the reconstruction problem at point $\mathbf{q}$ to the problem of computing the transformation between pairs of 1D functions, namely the integrals of the near and far reflectance images along the x- and (possibly) the y-axes. These 1D functions can be acquired by capturing images of the scene while displaying a single stripe on the monitor (*i.e.*, $I(x, y) = 1$ along a single row or column and 0 everywhere else). For an $N \times N$-pixel LCD, this requires capturing $4N$ images, one per row/column per source position.

$R_{\mathbf{q}}^x$ (blue curve) and $\hat{R}_{\mathbf{q}}^{\hat{x}}$ (red curve) $\mathcal{R}_{\mathbf{q}}^x$ (blue curve) and $\hat{\mathcal{R}}_{\mathbf{q}}^{\hat{x}}$ (red curve)
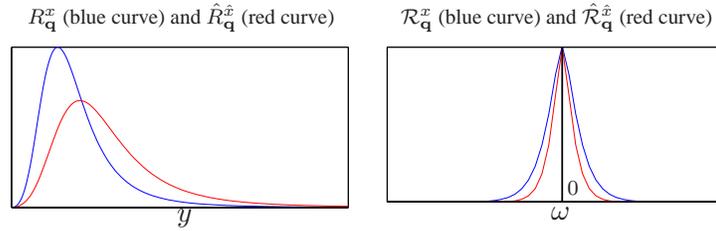
Figure 5.3: Reflectance magnification in the spatial and frequency domains. The plots correspond to the middle example in Figure 5.4, with $\sigma = 2$.

Despite the relative efficiency of this general procedure, it still involves capturing a potentially large number of samples of the functions $R_{\mathbf{q}}^x$ and $\hat{R}_{\mathbf{q}}^{\hat{x}}$, and thus a large number of images (each image gives us one sample). To reduce sampling requirements even further, we analyze reflectance magnification in the frequency domain. This allows us to take advantage of the broad spectral support of these functions and compute depth by sampling just a few of their Fourier coefficients.

5.3   Reflectance Magnification in the Frequency Domain

How does the spectrum of $R_{\mathbf{q}}^x$ change when the light source moves to the far plane? The following observation shows that as the source moves away from the scene, the spectrum shrinks linearly according to the source-relative depth (Figure 5.3):

Observation 3 (Reflectance Magnification in the Frequency Domain) *If $\mathcal{R}_{\mathbf{q}}^x$ and $\hat{\mathcal{R}}_{\mathbf{q}}^{\hat{x}}$ are the Fourier transforms of $R_{\mathbf{q}}^x$ and $\hat{R}_{\mathbf{q}}^{\hat{x}}$, respectively, then*

$$\left| \hat{\mathcal{R}}_{\mathbf{q}}^x [\omega] \right| = \left| \mathcal{R}_{\mathbf{q}}^x \left[\frac{\sigma + 1}{\sigma} \omega \right] \right| . \tag{5.11}$$
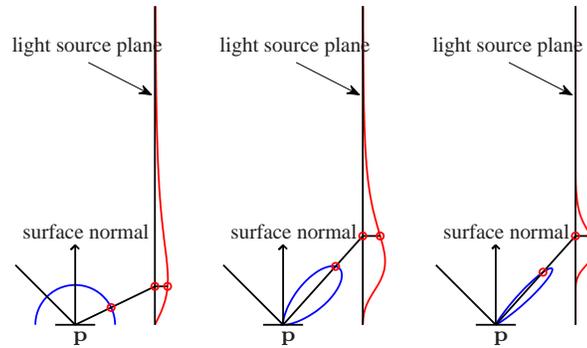
Figure 5.4: 1D slices of the reflectance function (red) for three different BRDFs (blue)—a diffuse BRDF on the left, and two BRDFs with an increasingly-narrow specular lobe. Note that the reflectance function quickly diminishes for points on the light source away from $\mathbf{p}$.

See the Appendix for a derivation of Eq. (5.11). The simple stretching relation between near and far spectra makes shape computation especially simple in the frequency domain. We use three basic ideas to do this.

First, reflectance images include the effects of foreshortening and squared-distance falloff (Eq. (5.3)) that cause significant spatial attenuation. As a result, both the images themselves and their line integrals have limited spatial support—and thus broad spectral support—regardless of a point's BRDF (Figure 5.4). This allows us to estimate spectral stretching from just a few Fourier coefficients in the vicinity of the DC component.

Second, computing a Fourier coefficient $\mathcal{R}_{\mathbf{q}}^{x}[\omega]$ can be done optically, by displaying the pattern $I(x, y) = \cos(2\pi\omega y) - j\sin(2\pi\omega y)$ on the light source plane and measuring the (complex-valued) radiance at $\mathbf{q}$. Since we cannot display and capture complex-valued signals

directly, we capture the coefficient in parts, using four patterns:[2]

$$\gamma_c^+ \text{ from } \max\{0, \cos(2\pi\omega y)\} \qquad\qquad \gamma_s^+ \text{ from } \max\{0, \sin(2\pi\omega y)\} \qquad (5.12)$$

$$\gamma_c^- \text{ from } \max\{0, -\cos(2\pi\omega y)\} \qquad\qquad \gamma_s^- \text{ from } \max\{0, -\sin(2\pi\omega y)\} \qquad (5.13)$$

$$\text{with } \mathcal{R}_\mathbf{q}^x [\omega] = (\gamma_c^+ - \gamma_c^-) - j(\gamma_s^+ - \gamma_s^-) \,. \qquad (5.14)$$

Third, since the line integrals $R_\mathbf{q}^x$ and $\hat{R}_\mathbf{q}^{\hat{x}}$ are typically just an isolated peak, and since this peak will occur at corresponding positions on the near and far plane,[3] it can be localized by temporal phase unwrapping from the available Fourier coefficients [19, 20, 53].

We turn these ideas into concrete algorithms below.

5.4 Depth from Reflectance Magnification

5.4.1 Source-Relative Depth for Differentiable BRDFs

We apply Eq. (5.11) directly to find the source-relative depth for a particular pixel. Specifically, each four-pattern sequence in Eqs. (5.12)-(5.14) provides a single Fourier coefficient, $\mathcal{R}_\mathbf{q}^x [\omega]$, and thus a single sample of the spectrum of $R_\mathbf{q}^x$. We repeat this acquisition procedure for a fixed set of n frequencies to obtain n samples of the spectrum, for each of the two light source positions ($8n$ images in total). To compute the stretch, we fit smoothing splines to $\mathcal{R}_\mathbf{q}^x [\omega]$ and $\hat{\mathcal{R}}_\mathbf{q}^x [\omega]$, to obtain $\mathcal{S}_\mathbf{q}^x [\omega]$ and $\hat{\mathcal{S}}_\mathbf{q}^x [\omega]$ respectively according to [101]. We then compute the following error function between the near spline and the stretched far plane spline:

---

[2]Intuitively, this can be thought of as a direct optical implementation of the Fourier transform. An analogous display procedure was used in [97, 139] for capturing environment mattes.

[3]Equation (5.10) implies that if a local extremum of $R_\mathbf{q}^x$ occurs at position y, then a local extremum of $\hat{R}_\mathbf{q}^x$ will occur at position $\hat{y}$, with $\hat{y}$ given by Eq. (5.7).

$$E(\sigma) = \sum_{\omega} (\mathcal{S}_{\mathbf{q}}^x[\omega] - \left(\frac{\sigma+1}{\sigma}\right) \hat{\mathcal{S}}_{\mathbf{q}}^x[\omega])^2 \tag{5.15}$$

We find the σ that minimizes E, giving the following algorithm:

Step 1 Capture near and far plane frequencies $\mathcal{R}^x(\omega)$ and $\hat{\mathcal{R}}^x(\omega)$ and $\mathcal{R}^y(\omega)$ and $\mathcal{R}^y(\omega)$

(Eqs. (5.12)-(5.14))

Step 2 For each pixel q:

Step 2.1 Fit smoothing splines to get $\mathcal{S}_{\mathbf{q}}^x[\omega]$ and $\hat{\mathcal{S}}_{\mathbf{q}}^x[\omega]$

Step 2.2 Perform search for σ that minimizes E

In practice we used $n = 10$ frequencies on the near screen, spanning the DC, $N/2$, $N/4$ ... $N/512$ and $\hat{n} = 5$ frequencies on the far screen spanning the DC, $N/2$, $N/4$ ... $N/16$, where N is the maximum spatial extent of the monitor screen. Due to the stretching, we required fewer samples on the far screen to match the near screen curve.

An important feature of this algorithm is that it is completely calibration free: it requires no knowledge of the intrinsic or extrinsic properties of the camera and requires no knowledge of the actual translation of the monitor (source-relative depths are recovered as fractions of this unknown distance).

5.4.2 Camera-Relative Depth for Differentiable and Specular BRDFs

In order to obtain camera-relative depth, we seek to compute one pair of corresponding points on the two light source planes ($\mathbf{l}_{xy}$ and $\hat{\mathbf{l}}_{\hat{x}\hat{y}}$ in Figure 5.2 (b)). We then intersect the ray passing

through them with the view ray through $\mathbf{q}$. This idea has been used repeatedly for reconstruction in the specular case [15, 69, 85] and here we extend this procedure to the non-specular cases as well.

In the case of a specular BRDF, the reflectance image will be a delta function on both planes. In real images, it will correspond to a sharp peak in $R_\mathbf{q}(x, y)$. We can compute these peaks by transforming the same frequencies captured in the previous section (Section 5.4.1) into the spatial domain using the inverse Fourier transform.

So even in the case of non-specular BRDFs the global extrema in the reflectance images also lie on the same 3D ray, we can apply the same principle for reconstruction here as well.

This leads to the following algorithm:

Step 1 Capture near and far plane frequencies $\mathcal{R}^x(\omega)$ and $\hat{\mathcal{R}}^x(\omega)$ and $\mathcal{R}^y(\omega)$ and $\mathcal{R}^y(\omega)$ (Eqs. (5.12)-(5.14))

Step 2 Set the remaining $(N-n)$ and $(N-\hat{n})$ frequencies to zero and compute approximations of R'_x, $\hat{R}'_x$ and R'_y, $\hat{R}'_y$ by the inverse Fourier transform

Step 3 For each pixel $\mathbf{q}$:

    Step 3.1 Find $\mathbf{l}_{xy}$ and $\hat{\mathbf{l}}_{\hat{x}\hat{y}}$ by taking the maxima of R'_x, R'_y and $\hat{R}'_x$, $\hat{R}'_y$.

    Step 3.2 Intersect this ray with viewing ray through $\mathbf{q}$ to obtain $\mathbf{p}$

Note that this does not require more frequencies than those used in Section 5.4.1, and is equally efficient. The only difference is the calibration requirement.

5.5 Experiments

In our experiments we used an 20 inch LCD screen as our illumination source and we translated it on a track perpendicular to the screen's plane by a distance of t (Figure 5.1). We placed the objects as close to the near plane P_n as possible while still being visible to the camera. This placement is important to minimize the number of incident rays from P_n that are not also covered by P_f since the planes are not infinite (Figure 5.1). We chose t based on the size of the target object, with deeper objects needing a larger t.

We corrected the sinusoidal patterns to take into account the LCD's gamma function and for each pattern we displayed two images one of the positive component and one of the negative component and later subtracted the negative component image as in [36]. The camera was radiometrically calibrated using the method described in [26] and for some of scenes we captured multiple exposures (usually only one or two was sufficient) and created HDR images for input to our algorithm [26]. For each exposure we required 60 images spanning 10 frequencies on P_n and 5 on P_f.

In the following sections we present reconstruction results from both the uncalibrated (Section 5.4.1) and calibrated (Section 5.4.2) methods. We show the recovered depth maps as well as the half-angle vector maps for each object and we compare to approximate ground truth from laser scans. The laser scan data was acquired with a Konica Minolta VI-9i at approximately the same distance as our camera to the scenes. We matched the laser scans to our reconstructed points by a least squares best-fit approach after manual initialization. For the specular datasets we had to coat the objects with white powder before scanning.

Orange Dataset This dataset is a scan of the skin of a tangerine orange whose material has both diffuse and specular components (Figure 5.5). We used an $t = 20mm$ for the scan and

two exposures. The uncalibrated depth recovery closely matches the laser scan data. The navel of the orange was not recovered since the signal was too dark at this point. The calibrated depth map shows the subtle wrinkles of the skin, while the wrinkles are less obvious in the laser scan. This may be because the sub-surface scattering in the laser scan causes artificial smoothing making our method more attractive, if slightly noisier (see the cross section in Figure 5.6).

The half angle vectors here vary significantly from the true normal map because the orange has a significant diffuse component in addition to the specular component.

Dome dataset This object is a white plaster dome on a pedestal created by a 3D printer and its reflectance is very diffuse. This object is a good test for our approach in terms of its reflectance as well as interesting geometry with the planar pedestal and the intersection with the smoothly curving dome.

We used two exposures for this scene, with $t = 40mm$ and the depths from P_n varied from $5mm$ to $85mm$. Notice that there are some errors due to indirect illumination, especially at the intersection of the dome with the pedestal. This dataset illustrates how the accuracy of the reconstruction degrades as the distance from the illumination source increases, especially when it approaches $2t$, see cross section 2 in Figure 5.8. Again the half angle vectors do not model the normal vectors well here because of the diffuse reflectance of the dome.

Sphere dataset We scanned a mirroring sphere with nearly pure specular reflectance using a single exposure time and $t = 20mm$. The laser scan required coating the sphere because of its high reflectivity. As expected the uncalibrated approach did not recover the depth well, however the calibrated method achieved good results (Figure 5.9). It is possible to see some ringing artifacts on the sphere, this is due to the low frequency of the displayed patterns. With a higher number, such artifact would be reduced but this is a trade-off with capture time. The half angle vectors in this case give very good estimates of the surface normals as we would

expect from a specular object.

Cylinder dataset We also scanned a highly specular cylinder (the top of a silver spray paint can) to further demonstrate our specular reconstruction. Again the laser scan required a coating on the cylinder. For our scan we used a single exposure time and $t = 20mm$. Similar ringing artifacts show up as in the sphere case, however the curvature of the cylinder is correctly recovered as can be seen in the mesh view and cross section of Figure 5.10. The half angle vectors again give very good estimates of the surface normals as we would expect.

Can-opener dataset Our final dataset consists of a can-opener lying on top of a towel. This scene has multiple materials: the glossy metal on the can-opener, the plastic handle is diffuse and also has significant sub-surface scattering and the towel is also diffuse with highly complex local geometry that causes inter-reflection and indirect light. Given these materials, this is a difficult scene to reconstruct for almost all 3D reconstruction methods including ours and the 'ground truth' laser scan. We used $t = 40mm$ for this scan with four exposures. Figure 5.11 shows the recovered depth maps along with the laser scans. The glossy region on the can-opener was partially recovered by the uncalibrated algorithm and the towel shows some problems due to the indirect light. The calibrated reconstruction is more robust to the varying materials and has a close correspondence with the laser scan. Note, however the region of the towel inside the loop of the can-opener that exhibits errors between the calibrated depth map and the laser depths. This is caused by partial occlusion of the light source by the can-opener. The half-angle vectors on the metal of the can-opener match the laser normal map relatively closely as expected.

5.6 Conclusion

We have presented a new approach for recovering the shape of general scenes from a single viewpoint using a moving planar light source (e.g. an LCD monitor). Additionally, when calibration is available 3D points can be reconstructed and further robustness to specular materials gained. We have shown our method to work under a variety of difficult scenes exhibiting specular and diffuse materials. We are able to recover depth for each of these materials using a convenient illumination and capture system.

Our method is limited to direct light and thus inter-reflections and significant subsurface scattering or refraction can cause errors. Since we limit the number of illumination patterns to a few low frequencies, some high frequency BRDFs or occlusions can cause errors in the depths. However more images could be used to reduce this effect in a trade-off with capture time.

In the future we would like to experiment with broader sampling across the Fourier spectrum, rather that just using low frequencies. Testing alternative bases might also improve our results or be more efficient for capture. In addition we would like to experiment to see how precise the translation of P_n to P_f needs to be to maintain reasonable results, because reducing the requirement of a translation stage would further simplify the experimental setup.

Depth Map

Src-Relative Depths

Laser Depth Map

Laser Src-Relative Depths

|Depths - Laser Depths|

|S.R. Depths - Laser Depths|
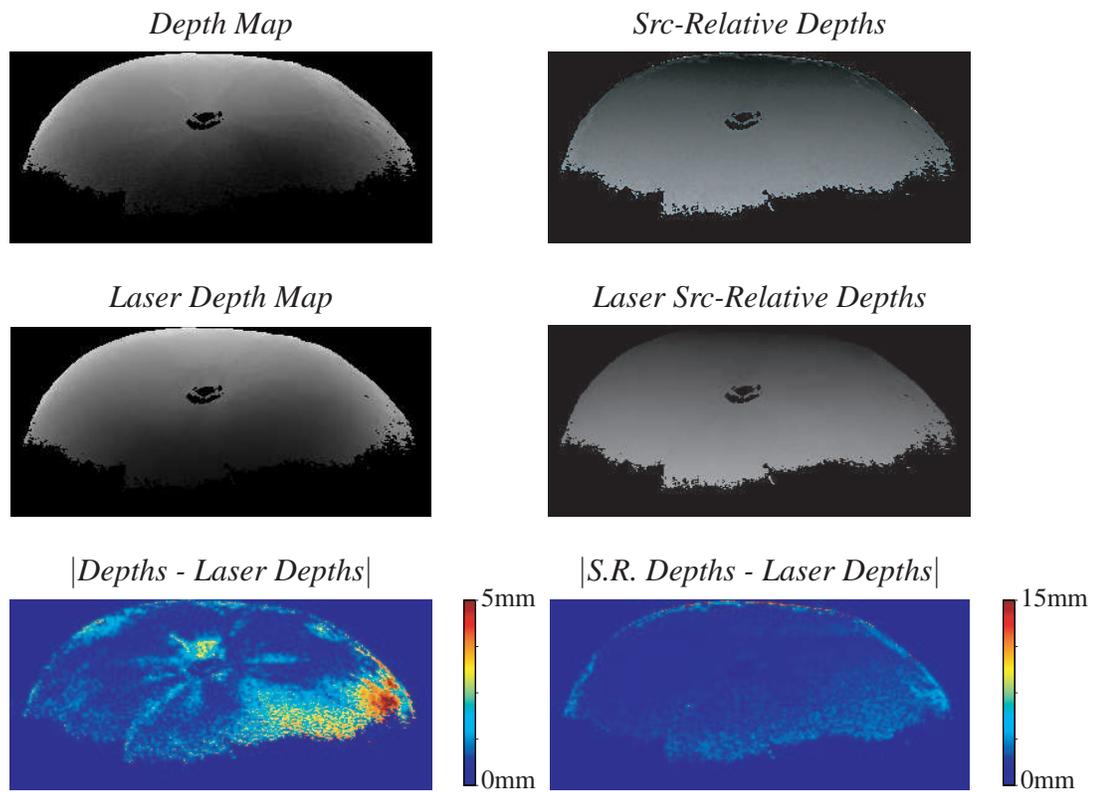
Figure 5.5: Orange dataset results

Half Vector Map

Laser Normal Map

|Half Vecs - Laser Normals|

Input Iimage

Cross Section

Reconstructed points
Laser points
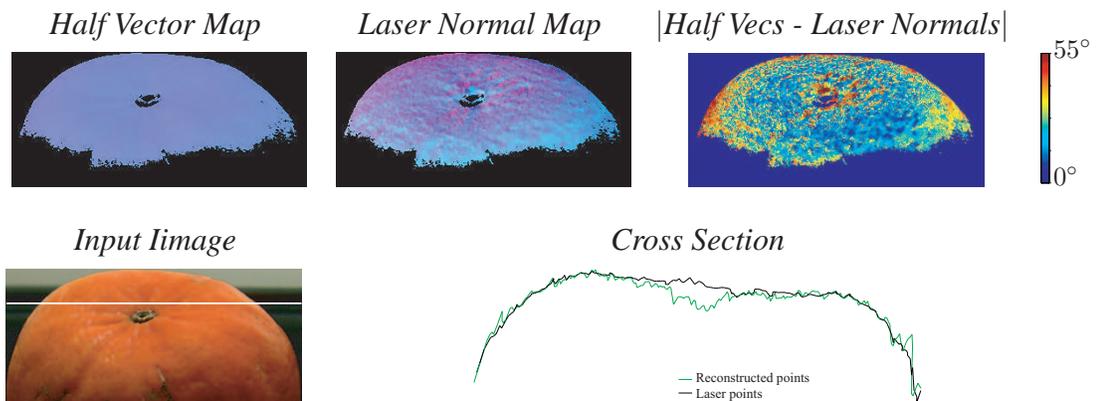
Figure 5.6: Orange dataset results

Depth Map

Src-Relative Depth

Laser Depth Map

Laser Src-Relative Depths

|Depths - Laser Depths|

|S.R. Depths - Laser Depths|
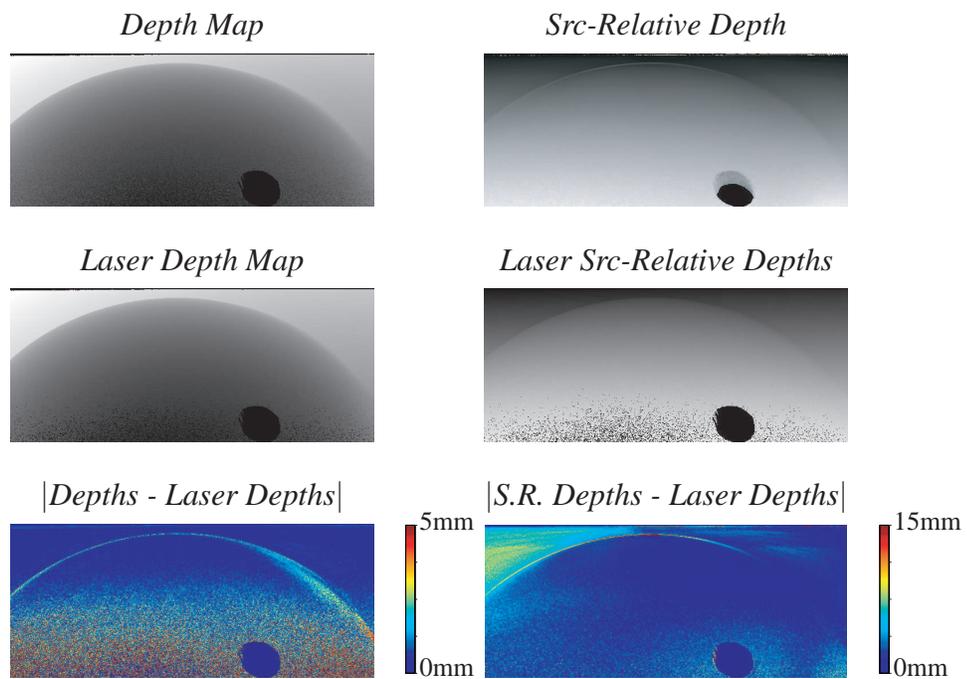
Figure 5.7: Dome dataset results

Half Vector Map   *Laser Normal Map*   *|Half Vecs - Laser Normals|*

Input Image   *Cross sections*

Reconstructed points
Laser points

Object Photo
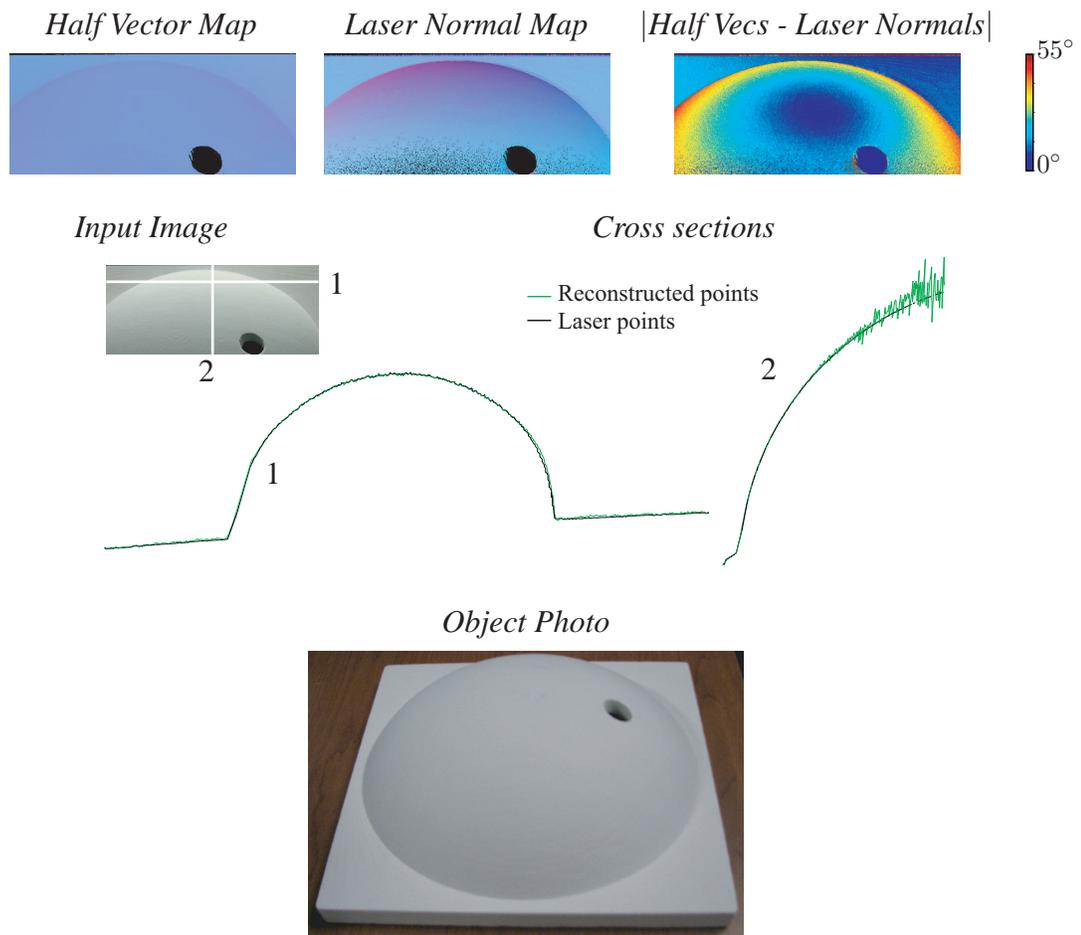
Figure 5.8: Dome dataset results

Src-Relative Depth   Laser Src-Relative Depths |S.R. Depths - Laser Depths|

Depth Map   Laser Depth Map |Depths - Laser Depths|

Half Vector Map   Laser Normal Map |Half Vecs - Laser Normals|

Input Image   Cross section

— Reconstructed points
— Laser points

Object Photo
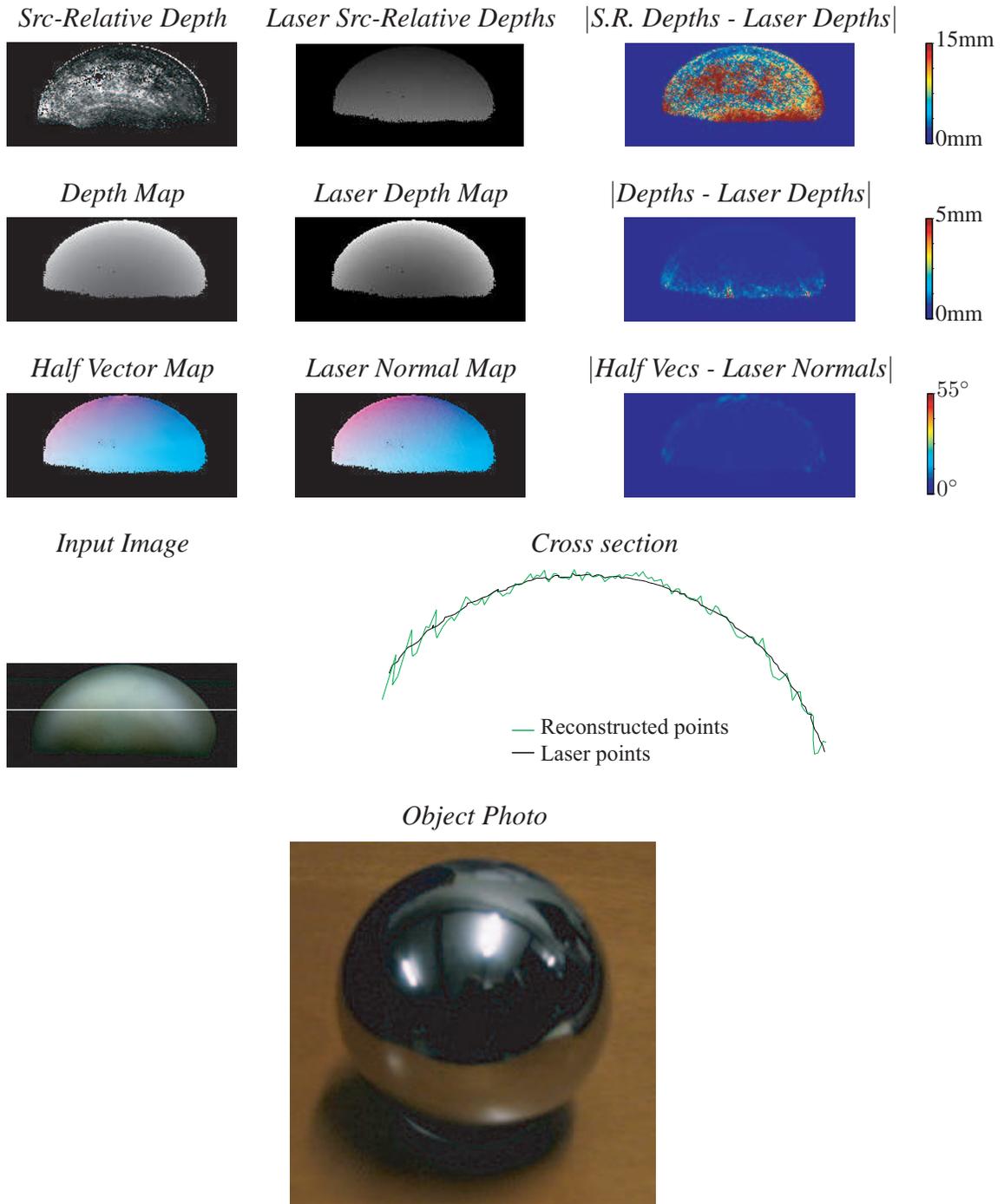
Figure 5.9: Sphere dataset results

Src-Relative Depths   Laser Src-Relative Depths |S.R. Depths - Laser Depths|

15mm

0mm

Depth Map   Laser Depth Map |Depths - Laser Depths|

5mm

0mm

Half Vector Map   Laser Normal Map |Half Vecs - Laser Normals|

55°

0°

Input Image   Cross section

—— Reconstructed points

—— Laser points

Object Photo
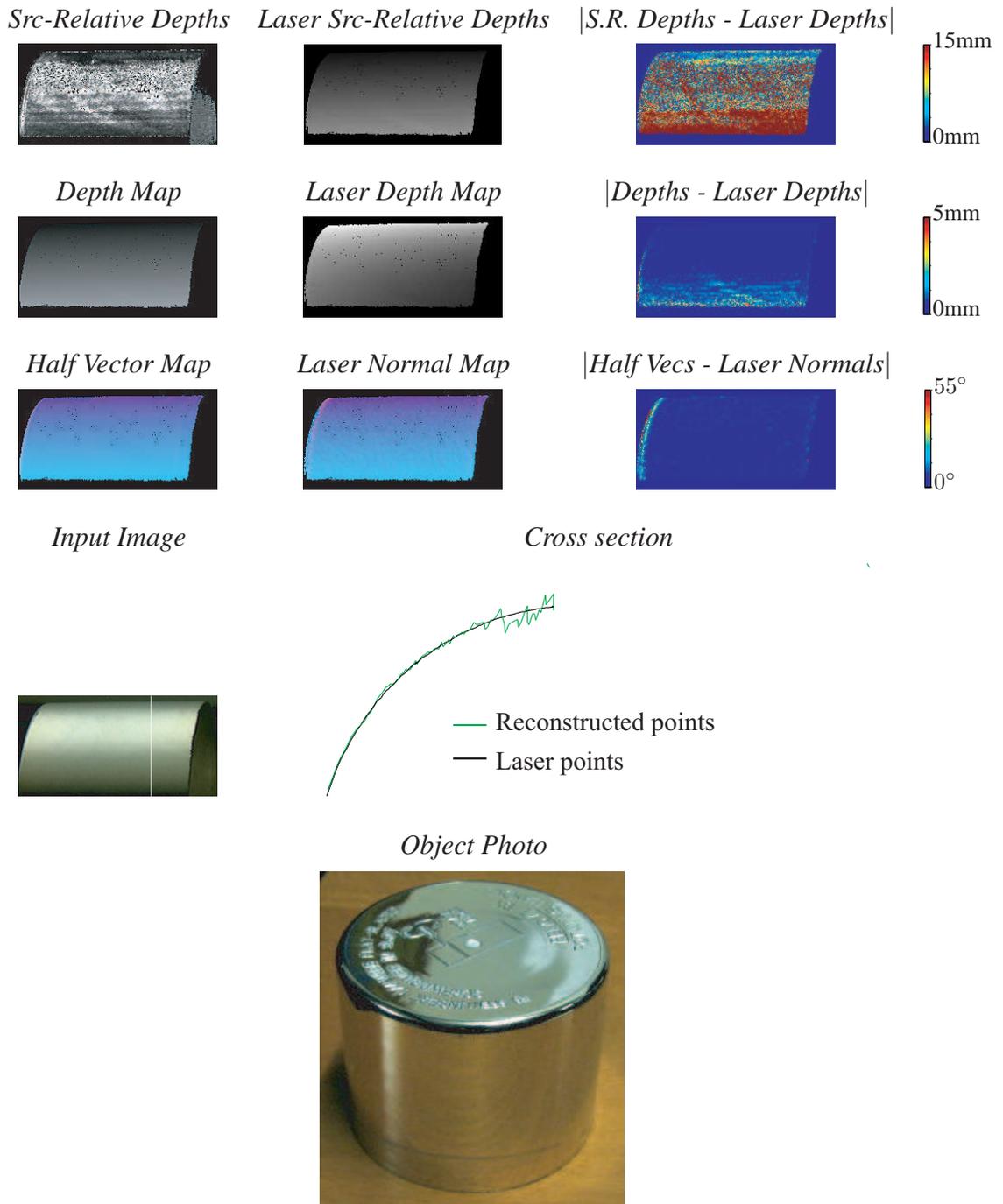
Figure 5.10: Cylinder dataset results

Src-Relative Depths

Laser Src-Relative Depths

|S.R. Depths - Laser Depths|

Depth Map

Laser Depth Map

|Depths - Laser Depths|
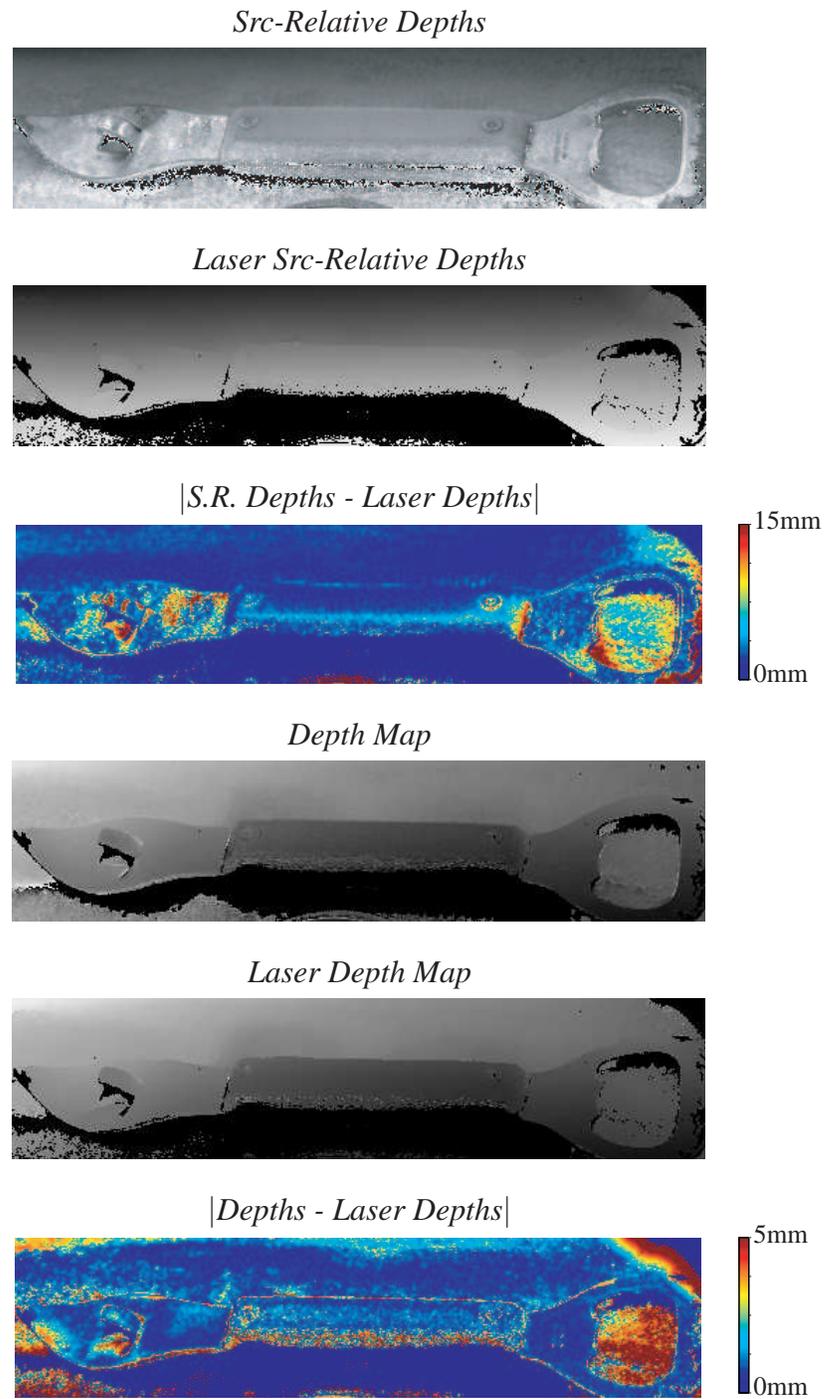

Figure 5.11: Can-opener dataset results
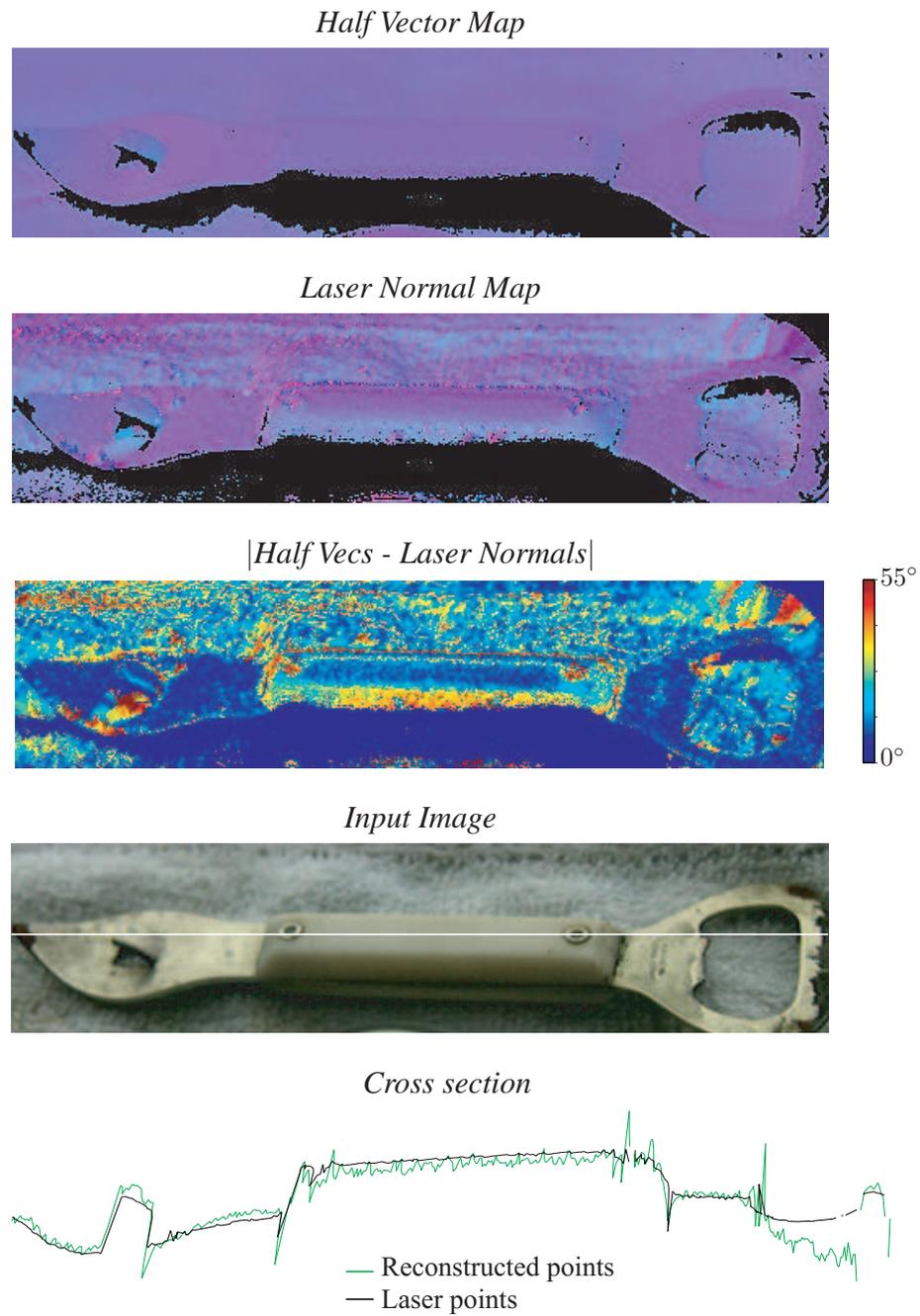
Figure 5.12: Can-opener dataset results

Chapter 6

Conclusions

"It is not by muscle, speed, or physical dexterity that great things are achieved, but by reflection, force of character, and judgment."

-Marcus Tullius Cicero

In recent years we are finally starting to see the emergence of 3D photography systems with high fidelity that are robust to the kind of complex optics we see every day. Scenes with complex refractive and reflective properties that were once thought to be impossible to scan are now coming within reach.

We have shown how simple multi-view camera systems as well as the utilization of multi-planar illumination sources can remove many of the ambiguities inherent in the optical complexity our target scenes. In particular we have shown that:

- We can recover 3D points and normals of temporally dynamic refractive surfaces using a simple stereo system using a known planar submerged pattern.

- We are able to reconstruct 3D points and normals from complex, inhomogenous trans-

parent objects with a multi-view system by observing "scatter-traces" at each pixel.

- We can reconstruct source-relative depth of scenes with a wide variety of BRDFs from a single uncalibrated view, and upon calibration we can recover 3D points on the surface and be more robust to truly specular materials.

While the main goal of this work has been to show how the first steps of tackling such problems with 3D photography, we have also begun to show how these techniques can be improved with more rapid capture as well as broadening the range of material types that can be scanned with one capture. It is not hard to see such systems being implemented in important industrial scanning tasks. All of these approaches require only simple camera and illumination systems and can readily be applied to 3D reconstruction tasks in artifact archival, oceanography, medical instrumentation, manufacturing quality assurance and possibly in robotic systems.

6.1 Future work

We see great potential in extending the work described here. In many cases we have just 'scratched' the surface in terms of what can be achieved.

Interior reconstruction. Our approach to 3D reconstruction of inhomogeneous transparent objects focuses on the exterior surface, leaving the interior largely unknown despite a wealth of information in the scatter trace. Given an exterior reconstruction and the refractive index, some initial experiments have shown that it is possible to recover opaque interiors through a modified version of multi-view stereo. This alone prompts interesting problems: how do we find the 3D pose of a transparent object relative to images of it, and how can we solve for the refractive index of such an object from photographs. It may also be possible to jointly solve

for the interior and exterior under some conditions. In addition, we would like to be able to recover transparent objects with non-opaque interior structure as well.

Material and scattering capture. While our initial work has focused on recovering the 3D shape of target objects, the wealth of information in scatter traces could enable us to simultaneously acquire BRDF models for materials on the object. If enough views are available or sufficient normals can be classified as belonging to the same material a dense sampling of the BRDF is possible. It would also be interesting to factor in scattering properties of a medium. It may be possible to estimate this from the indirect component of the scatter traces. Capturing the material and scattering properties would also enable more accurate rendering of scanned objects.

Exploring the spectrum. In the work presented here we assume monochromatic light, however refraction changes its properties according to wavelength. This effect could possible be used to estimate the refractive index or provide additional cues for depth, making our approach more efficient. Additionally it would be interesting to make use of infrared illumination as its transmission is significantly reduced compared to shorter wavelengths.

Appendix A

Proof of Ambiguity Theorem

Proof of Theorem 1 The proof uses two basic intuitions. First, given an arbitrary value for the refractive index, each viewpoint can be thought of as defining a *3D constraint curve*, representing all assignments of distances and normals to a pixel $\mathbf{q}$ that are compatible with Snell's law. Hence, an assignment that is consistent with both viewpoints corresponds to the intersection of two such constraint curves. Second, and most important, for an arbitrarily-shaped surface, these 3D curves will be in general position with respect to each other and, therefore, will not have a common intersection. From these two facts we conclude that when the refractive index has an arbitrary value, there will be no distance and normal assignment that is consistent with both viewpoints. Hence, such consistency can only be achieved for isolated refractive index values. We formalize these intuitions below.

Let r^* be the true refractive index of the surface and let $r \neq r^*$ be an arbitrary value of this index. Without loss of generality, we assume that the function $\mathbf{C}(\mathbf{q})$ is known for all pixels $\mathbf{q}$ and is continuous. Let $\mathbf{q}$ be an arbitrary pixel in the first viewpoint and let d^* be its true distance to the surface. Given value r for the refractive index, every distance d defines a unique

normal, $\mathbf{n}(d,r)$, compatible with Snell's law (Figure 3.2 and Eq. (3.2)). Suppose we represent unit vectors with two angles: an angle θ, corresponding to the angle between the vector and the ray through $\mathbf{q}$; and an angle ϕ, corresponding to the angle between the vector and the normal of $\mathbf{q}$'s refraction plane. In this representation, the distance and normal assignments to $\mathbf{q}$ that are compatible with Snell's law define a curve γ in (d, θ, ϕ)-space. This curve will always lie on the plane $\phi = 0$ since, by definition, the normal $\mathbf{n}(d,r)$ always lies on the refraction plane of pixel $\mathbf{q}$.

Now let $\mathbf{q}'(d)$ be the projection of $\mathbf{p}(d)$ in the second viewpoint (Figure 3.1). Since $\mathbf{C}(\mathbf{q}'(d))$ is known, there is only one normal, $\mathbf{n}'(d,r)$, that can be assigned to $\mathbf{p}(d)$ and is compatible with Snell's law in the second viewpoint. This normal will lie on the refraction plane of pixel $\mathbf{q}'$. Generically, the refraction planes of pixels $\mathbf{q}$ and $\mathbf{q}'$ are distinct. Hence, the normal $\mathbf{n}'(d,r)$ may not lie on the refraction plane of pixel $\mathbf{q}$ and, as d varies, $\mathbf{n}'(d,r)$ will trace a general curve γ' in (d, θ, ϕ)-space, i.e., a curve that is not restricted to the plane $\phi = 0$.

We now show that γ and γ' do not intersect. First note that the two curves cannot intersect in the neighborhood of the "true" distance d^* because $\mathbf{n}'(d^*, r) \neq \mathbf{n}(d^*, r)$.[1]

Now consider distances away from d^*. We show that γ' and γ generically will not intersect there either. In particular, the normal $\mathbf{n}'(d,r)$ is completely determined by point $\mathbf{C}(\mathbf{q}'(d))$ which, in turn, is determined by the normal of the true surface point projecting to pixel $\mathbf{q}'(d)$. Since $\mathbf{q}'(d)$ lies on the epipolar line of $\mathbf{q}$ for all values of d, it follows that curve γ' is completely determined by the surface normal of points at the intersection, C, of this epipolar plane with the true surface. For γ' and γ to intersect there must be a point on C outside the neighborhood of $\mathbf{p}(d^*)$ whose surface normal is identical to $\mathbf{p}(d^*)$. This condition, however, cannot be satisfied

---

[1] Observe that $\mathbf{n}'(d^*, r^*) = \mathbf{n}(d^*, r^*)$ since Snell's law is satisfied for both viewpoints in the true scene. Now, since there is a 1-1 correspondence between refractive indices and normals when d^* is fixed, and since the refraction planes of $\mathbf{q}$ and $\mathbf{q}'(d)$ have only one normal in common, it follows that $\mathbf{n}'(d^*, r) \neq \mathbf{n}(d^*, r)$.

for an open 2D set of points on a generic surface. It follows that γ' and γ are non-intersecting for almost all points on the surface and, hence, for almost all pixels in the surface's projection.

Appendix B

2D Scatter Traces

Suppose we illuminate the scene with a linear light source that is oriented along the z axis and can move on the xy-plane. The radiance incident at pixel $\mathbf{q}$ for light source position (x, y) is just the integral of the pixel's 3D scatter traces along z:

$$T_{\mathbf{q}}(x, y) = \int T_{\mathbf{q}}(x, y, z)dz \ . \tag{B.1}$$

Now consider each of the three types of light transport discussed in Section 4.2.1 for the 3D case. If the point projecting to $\mathbf{q}$ is a planar mirror, the above integral is just the 2D analog of the 3D impulse scatter trace: it is equal to the orthographic projection of $\mathbf{q}$'s 3D impulse scatter trace onto the xy-plane. As such, (1) its value decreases monotonically with increasing distance from the point of reflection, and (2) it is non-zero along a single 2D ray that corresponds to the direction of specular reflection, projected onto the xy-plane. Analysis of the other cases follows as a direct consequence of these two observations.

Appendix C

Reflectance Magnification Derivations

Derivation of Eq. (5.10):   Using Eq. (5.6) and integrating over $\hat{x}$ we have

$$\hat{R}_{\mathbf{q}}^x\left(\hat{y}\right) \;=\; \left(\frac{\sigma}{\sigma+1}\right)^2 \int_{\hat{x}} R_{\mathbf{q}}\left(x, y\right)\, d\hat{x} \tag{C.1}$$

$$=\; \left(\frac{\sigma}{\sigma+1}\right)^2 \int_{\hat{x}} R_{\mathbf{q}}\left(\frac{\sigma}{\sigma+1}\,\hat{x} + \frac{1}{\sigma+1}\,a,\; y\right)\, d\hat{x} \tag{C.2}$$

$$=\; \left(\frac{\sigma}{\sigma+1}\right) \int_{z} R_{\mathbf{q}}\left(z,\; y\right)\, dz \tag{C.3}$$

$$=\; \left(\frac{\sigma}{\sigma+1}\right) R_{\mathbf{q}}^x\left(y\right)\;,$$

where we used Eq. (5.7) to express x in terms of $\hat{x}$ in Eq. (C.2).   ∎

Derivation of Eq. (5.11)**:**

$$
\hat{\mathcal{R}}_{\mathbf{q}}^{\hat{x}}\left[\omega\right] = \int \hat{R}_{\mathbf{q}}^{\hat{x}}\left(\hat{y}\right)\exp\left(-j2\pi\omega\hat{y}\right)\ d\hat{y} \tag{C.4}
$$

$$
= \left(\frac{\sigma}{\sigma+1}\right)\int R_{\mathbf{q}}^{x}\left(y\right)\exp\left(-j2\pi\omega\hat{y}\right)\ d\hat{y} \tag{C.5}
$$

$$
= \left(\frac{\sigma}{\sigma+1}\right)\int R_{\mathbf{q}}^{x}\left(\frac{\sigma}{\sigma+1}\ \hat{y} + \frac{1}{\sigma+1}\ b\right)\exp\left(-j2\pi\omega\hat{y}\right)\ d\hat{y} \tag{C.6}
$$

$$
= \int R_{\mathbf{q}}^{x}\left(z\right)\exp\left(-j2\pi\frac{\sigma+1}{\sigma}\omega z\right)\exp\left(-j2\pi\omega\frac{b}{\sigma}\right)\ dz \tag{C.7}
$$

$$
= \exp\left(-j2\pi\omega\frac{b}{\sigma}\right)\mathcal{R}_{\mathbf{q}}^{x}\left[\frac{\sigma+1}{\sigma}\omega\right]\ , \tag{C.8}
$$

where we used Eq. (5.10) to obtain Eq. (C.5), and used Eq. (5.7) to express y in terms of $\hat{y}$ in Eq. (C.6).

Equation (5.11) now follows by taking the magnitude on both sides of Eq. (C.8). ■

Bibliography

[1] Y. Adato, Y. Vasilyev, O. Ben-Shahar, and T. Zickler. Toward a theory of shape from specular flow. In *Proc. IEEE 11th International Conference on Computer Vision ICCV 2007*, pages 1–8, 2007.

[2] E. H. Adelson and J. R. Bergen. The plenoptic function and the elements of early vision. In M. Landy and J. A. Movshon, editors, *Computation Models of Visual Processing*. MIT Press, 1991.

[3] Sameer Agarwal, Satya P. Mallick, David Kriegman, and Serge Belongie. On refractive optical flow. In *Proc. 8th European Conf. on Computer Vision*, pages 483–494, 2004.

[4] N. Alldrin, T. Zickler, and D. Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition CVPR 2008*, pages 1–8, 2008.

[5] N.G. Alldrin and D.J. Kriegman. Toward reconstructing surfaces with arbitrary isotropic reflectance : A stratified photometric stereo approach. In *Proc. IEEE 11th International Conference on Computer Vision ICCV 2007*, pages 1–8, 2007.

[6] Bradley Atcheson, Ivo Ihrke, Wolfgang Heidrich, Art Tevs, Derek Bradley, Marcus Magnor, and Hans-Peter Seidel. Time-resolved 3d capture of non-stationary gas flows. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 27(5):132, 2008.

[7] Mitsuru Baba, Kozo Ohtani, and Makoto Imai. New laser rangefinder for three-dimensional shape measurement of specular objects. *Optical Engineering*, 40(1):53–60, 2001.

[8] Simon Baker and Iain Matthews. Lucas-kanade 20 years on: A unifying framework. *IJCV*, 56(3):221–255, 2004.

[9] Moshe Ben-Ezra and Shree Nayar. What does motion reveal about transparency? In *Proc. 9th Int. Conf. on Computer Vision*, pages 1025–1032, 2003.

[10] J. Beynon. *Introductory University Optics*. Prentice Hall, 1996.

[11] D.N. Bhat and S.K. Nayar. Stereo and specular reflection. *IJCV*, 26(2):91–106, February 1998.

[12] A. Blake. Specular stereo. In *Int. Joint Conf. on Artificial Intelligence*, pages 973–976, 1985.

[13] A. Blake and G. Brelstaff. Geometry from specularities. In *Proc. 2nd Int. Conf. on Computer Vision*, pages 394–403, 1988.

[14] Jeremy De Bonet and Paul Viola. Roxels: Responsibility-weighted 3D volume reconstruction. In *Prof. 7th. Int. Conf. on Computer Vision*, pages 418–425, 1999.

[15] Thomas Bonfort and Peter Sturm. Voxel carving for specular surfaces. In *Proc. 9th Int. Conf. on Computer Vision*, pages 591–596, 2003.

[16] Thomas Bonfort, Peter Sturm, and Pau Gargallo. General specular surface triangulation. In *Proc. Asian Conference on Computer Vision*, pages 872–881, 2006.

[17] Jean-Yves Bouguet. MATLAB camera calibration toolbox. "http://www.vision.caltech.edu/bouguetj/calib_doc/".

[18] Tongbo Chen, Michael Goesele, and Hans-Peter Seidel. Mesostructure from specularity. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006)*, volume 2, pages 1825–1832, New York, NY, USA, 2006. IEEE.

[19] Tongbo Chen, H.P.A. Lensch, C. Fuchs, and H.-P. Seidel. Polarization and phase-shifting for 3d scanning of translucent objects. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition CVPR '07*, pages 1–8, 17–22 June 2007.

[20] Tongbo Chen, Hans-Peter Seidel, and Hendrik P. A. Lensch. Modulated phase-shifting for 3d scanning. In *CVPR*, 2008.

[21] Yung-Yu Chuang, Douglas E. Zonkger, Joel Hirdorff, Brian Curless, and David H. Salesin. Environment matting extensions: Toward higher accuracy and real-time capture. In *Proc. ACM SIGGRAPH*, pages 121–130, 2000.

[22] Hin-Shun Chung and Jiaya Jia. Efficient photometric stereo on glossy surfaces with wide specular lobes. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition CVPR 2008*, 2008.

[23] J.M. Daida, D. Lund, C. Wolf, G.A. Meadows, K. Schroeder, J. Vesecky, D.R. Lyzenga, B.C. Hannan, and R.R. Bertram. Measuring topography of small-scale water surface waves. In *Geoscience and Remote Sensing Symposium. Conference Proceedings*, volume 3, pages 1881–1883, 1995.

[24] James E. Davis, Ruigang Yang, and Liang Wang. Brdf invariant stereo using light transport constancy. In *Proc. 10th Int. Conf. Computer Vision*, pages 436–443, 2005.

[25] Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. Acquiring the reflectance field of a human face. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 145–156, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.

[26] Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *Proc. ACM SIGGRAPH*, pages 369–378, 1997.

[27] Yuanyuan Ding and Jingyi Yu. Recovering shape characteristics on near-flat specular surfaces. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition CVPR 2008*, pages 1–8, 2008.

[28] Yuanyuan Ding, Jingyi Yu, and Peter Sturm. Recovering specular surfaces using curved line images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Miami Beach, Florida, June 2009.

[29] Dave Edwards, Solomon Boulos, Jared Johnson, Peter Shirley, Michael Ashikhmin, Michael Stark, and Chris Wyman. The halway vector disk for brdf modeling. *ACM Trans. on Graphics*, 25(1):1–18, 2006.

[30] Douglas Enright, Stephen Marschner, and Ronald Fedkiw. Animation and rendering of complex water surfaces. In *Proc. ACM SIGGRAPH*, pages 736–744, 2002.

[31] Phillipp Flach and Hans-Gerd Maas. Vision-based techniques for refraction analysis in applications of terrestial geodesy. In *Int. Archives of Photogrammetry and Remote Sensing*, pages 195–201, 2000.

[32] Roland W Fleming, Antonio Torralba, and Edward H Adelson. Specular reflections and the perception of shape. *J Vis*, 4(9):798–820, Sep 2004.

[33] Y. Francken, T. Cuypers, T. Mertens, J. Gielis, and P. Bekaert. High quality mesostructure acquisition using specularities. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition CVPR 2008*, pages 1–7, 23–28 June 2008.

[34] Andrew S. Glassner. *Principles of Digital Image Synthesis*. Morgan Kaufmann, 1995.

[35] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S.M. Seitz. Multi-view stereo for community photo collections. In *Proc. IEEE 11th International Conference on Computer Vision ICCV 2007*, pages 1–8, 14–21 Oct. 2007.

[36] Michael Goesele, Xavier Granier, Wolfgang Heidrich, and Hans-Peter Seidel. Accurate light source acquisition and rendering. *ACM Trans. Graph.*, 22(3):621–630, July 2003.

[37] Michael Goesele, Hendrik P. A. Lensch, Jochen Lang, Christian Fuchs, and Hans-Peter Seidel. Disco: acquisition of translucent objects. *ACM Trans. Graph.*, 23(3):835–844, 2004.

[38] B. Goldlucke, I. Ihrke, C. Linz, and M. Magnor. Weighted minimal hypersurface reconstruction. *IEEE Trans. Pattern Anal. Machine Intell.*, 29(7):1194–1208, 2007.

[39] D.B. Goldman, B. Curless, A. Hertzmann, and S.M. Seitz. Shape and spatially-varying brdfs from photometric stereo. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 341–348Vol.1, 17-21 Oct. 2005.

[40] Mark A. Halstead, Brian A. Barsky, Stanley A. Klein, and Robert B. Mandell. Reconstructing curved surfaces from specular reflection patterns using spline surface fitting of normals. In *Proc. ACM SIGGRAPH*, pages 335–342, 1996.

[41] F.H. Harlow and J.E. Welch. Numerical calculation of time-dependent viscous incompressible flow. *Phys. Fluids*, 8:2182–2189, 1965.

[42] C. Harris and M. Stephens. A combined edge and corner detector. In *Proc. of the 4th Alvey Vision Conference*, pages 189–192, 1988.

[43] R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[44] Richard I. Hartley and Peter Sturm. Triangulation. *Computer Vision and Image Understanding: CVIU*, 68(2):146–157, 1997.

[45] Aaron Hertzman and Steven M. Seitz. Shape and materials by example: A photometric stereo approach. In *Proc. Computer Vision and Pattern Recognition Conf.*, pages 533–540, 2003.

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

[47] J. Höhle. Reconstruction of the underwater object. *Photogrammetric Engineering*, pages 948–954, 1971.

[48] Michael Holroyd, Jason Lawrence, Greg Humphreys, and Todd Zickler. A photometric approach for estimating normals and tangents. *ACM Transactions on Graphics (Proce. SIGGRAPH Asia 2008)*, 27(5):133, 2008.

[49] Hugues Hoppe, Tony Derose, Tom Duchamp, Mark Halstead, Hubert Jin, John McDonald, Jean Schweitzer, and Werner Stuetzle. Piecewise smooth surface reconstruction. In *ACM SIGGRAPH 1994 Conference Proceedings*, pages 295–302, 1994.

[50] B. K. P. Horn. *Robot Vision*. MIT Press, 1986.

[51] Berthold K. P. Horn and Michael J. Brooks. Shape from shading. MIT Press, 1989.

[52] Matthias B. Hullin, Martin Fuchs, Ivo Ihrke, Hans-Peter Seidel, and Hendrik P. A. Lensch. Fluorescent immersion range scanning. *ACM Transactions on Graphics*, 27(3):87:1–87:10, August 2008.

[53] J. M. Huntley and H. Saldner. Temporal phase-unwrapping algorithm for automated interferogram analysis. *Appl. Opt.*, 32(17):3047–3052, 1993.

[54] I. Ihrke, B. Goidluecke, and M. Magnor. Reconstructing the geometry of flowing water. In *Proc. Tenth IEEE International Conference on Computer Vision ICCV 2005*, volume 2, pages 1055–1060 Vol. 2, 2005.

[55] Katsushi Ikeuchi. Determining the surface orientations of specular surfaces by using the photometric stereo method. *IEEE Trans. Pattern Anal. Machine Intell.*, 3(6):661–669, 1981.

[56] Katsushi Ikeuchi and B. K. P. Horn. Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, 17:141–184, 1981.

[57] Guy A. Orban J. Farley Norman, James T. Todd. Perception of three-dimensional shape from specular highlights, deformations of shading, and other types of visual information. *Psychological Science*, 15(8):565, 2004.

[58] Bernd Jähne, Jochen Klinke, Peter Geissler, and Frank Hering. Image sequence analysis of ocean wind waves. In *Proc. of the International Seminar on Imaging in Transport Processes*, 1992.

[59] Bernd Jähne, Jochen Klinke, and Stefan Waas. Imaging of short ocean wind waves: a critical theoretical review. *J. Opt. Soc. Am. A.*, 11(8):2197–2209, 1994.

[60] Philipp Jenke, Michael Wand, Martin Bokeloh, Andreas Schilling, and Wolfgang Straßer. Bayesian point cloud reconstruction. *Computer Graphics forum*, 25(3):379–388, September 2006.

[61] H. Jin, S. Soatto, and A. J. Yezzi. Multi-view stereo beyond lambert. In *Proc. Computer Vision and Pattern Recognition Conf.*, volume 1, pages 171–178, 2003.

[62] Jürgen Kaminski, Svenja Lowitzsch, Markus C. Knauer, and Gerd Häusler. Full-field shape measurement of specular surfaces. In *The 5th International Workshop on Automatic Processing of Fringe Patterns*, 2005.

[63] Michael Kazhdan, Matthew Bolitho, and Hugues Hoppe. Poisson surface reconstruction. In *Eurographics Symposium on Geometry Processing*, 2006.

[64] W. C. Keller and B. L. Gotwols. Two-dimensional optical measurement of wave slope. *Applied Optics*, 22(22):3476–3478, 1983.

[65] B.B. Kimia and K. Siddiqi. Geometric heat equation and nonlinear diffusion of shapes and images. In K. Siddiqi, editor, *Proc. CVPR '94. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 113–120, 1994.

[66] Jan J. Koendering and A. J. van Doorn. The structure of two-dimensional scalar fields with applications to vision. *Biological Cybernetics*, 33:151–158, 1979.

[67] Sanjeev J. Koppal and Srinivasa G. Narasimhan. Clustering appearance for scene analysis. In *Proc. Computer Vision and Pattern Recognition Conf.*, pages 1003–1012, 2006.

[68] S.J. Koppal and S.G. Narasimhan. Novel depth cues from uncalibrated near-field lighting. In *Proc. IEEE 11th International Conference on Computer Vision ICCV 2007*, pages 1–8, 14–21 Oct. 2007.

[69] Kiriakos N. Kutulakos and Eron Steger. A theory of and refractive and specular 3d shape by light-path triangulation. In *Proc. 10th Int. Conf. on Computer Vision*, pages 1448–1455, 2005.

[70] Kiriakos N. Kutulakos and Eron Steger. A theory of specular and refractive shape by light-path triangulation. Technical Report MSR-TR-2005-12, Microsoft Research, 2005.

[71] Johann Heinrich Lambert. *Photometria sive de mensura et gradibus luminis, colorum et umbrae*. Eberhard Klett, 1760.

[72] Hendrik Peter Asmus Lensch. *Efficient, Image-Based Appearance Acquisition of Real-World Objects*. PhD thesis, Universitat des Saarlandes, 2003.

[73] Anat Levin, Asaf Zomet, and Yair Weiss. Separating reflections from a single image using local features. In *Proc. Computer Vision and Pattern Recognition Conf.*, pages 306–313, 2004.

[74] Marc Levoy and Pat Hanrahan. Light field rendering. In *Proc. ACM SIGGRAPH*, pages 31–42, 1996.

[75] S. Lin, Y. Lee, S. Kang, X. Tong, and H.-Y. Shum. Diffuse-specular separation and depth recovery from image sequences. In *Proc. of the European Conference on Computer Vision (ECCV)*, 2002.

[76] Frank Losasso, Frédéric Gibou, and Ron Fedkiw. Simulating water and smoke with an octree data structure. *ACM Trans. Graph.*, 23(3):457–462, 2004.

[77] Wan-Chun Ma, Tim Hawkins, Pieter Peers, Charles-Felix Chabert, Malte Weiss, and Paul Debevec. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In *Eurographics Symposium on Rendering 2007*, 2007.

[78] Hans-Gerd Maas. New developments in multimedia photogrammetry. In *Optical 3D Measurement Techniques III*. Wichmann Verlag, 1995.

[79] S. Magda, D. Kriegman, T. Zickler, and P. Belhumeur. Beyond Lambert: Reconstructing surfaces with arbitrary BRDFs. In *Proc. 8th Int. Conf. on Computer Vision*, pages 391–399, 2001.

[80] S.P. Mallick, T.E. Zickler, D.J. Kriegman, and P.N. Belhumeur. Beyond lambert: reconstructing specular surfaces using color. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 619–626vol.2, 20-25 June 2005.

[81] Stephen Robert Marschner. *INVERSE RENDERING FOR COMPUTER GRAPHICS*. PhD thesis, Cornell University, 1998.

[82] Wojciech Matusik, Hanspeter Pfister, Remo Ziegler, Addy Ngan, and Leonard McMillan. Acquisition and rendering of transparent and refractive objects. In *Proc. 13th Eurographics Workshop on Rendering*, pages 267–278, 2002.

[83] Daisuke Miyazaki and Katsushi Ikeuchi. Inverse polarization raytracing: Estimating surface shapes of transparent objects. In *Proc. Computer Vision and Pattern Recognition Conf.*, pages 910–917, 2005.

[84] Nigel Morris and Kiriakos N. Kutulakos. Dynamic refraction stereo. In *Proc. 10th Int. Conf. Computer Vision*, pages 1573–1580, 2005.

[85] Nigel J. W. Morris and Kiriakos N. Kutulakos. Reconstructing the surface of inhomogeneous transparent scenes by scatter-trace photography. In *Proc. IEEE 11th Int. Conf. on Computer Vision ICCV 2007*, 2007.

[86] Matthias Müller, David Charypar, and Markus Gross. Particle-based fluid simulation for interactive applications. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 154–159, 2003.

[87] H. Murase. Surface shape reconstruction of an undulating transparent object. In *Proc. 3rd Int. Conf. on Computer Vision*, pages 313–317, 1990.

[88] Shree K. Nayar, Katsushi Ikeuchi, and Takeo Kanade. Surface reflection: Physical and geometrical perspectives. *IEEE Trans. Pattern Anal. Machine Intell.*, 13(7):611–634, 1991.

[89] Shree K. Nayar, Gurunandan Krishnan, Michael D. Grossberg, and Ramesh Raskar. Fast separation of direct and global components of a scene using high frequency illumination. In *Proc. ACM SIGGRAPH*, pages 935–944, 2006.

[90] S.K. Nayar, K. Ikeuchi, and T. Kanade. Shape from interreflections. In *Computer Vision, 1990. Proceedings, Third International Conference on*, pages 2–11, 4-7 Dec. 1990.

[91] Diego Nehab, Szymon Rusinkiewicz, James Davis, and Ravi Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. In *Proc. ACM SIGGRAPH*, pages 536–543, 2005.

[92] F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis. Geometric considerations and nomenclature for reflectance. *National Bureau of Standards*, 1977.

[93] Atsushi Okamoto. Orientation problem of two-media photographs with curved boundary sur-
faces. *Photogrammetric Engineerning and Remote Sensing*, pages 303–316, 1984.

[94] Thanda Oo, Hiroshi Kawasaki, Yutaka Ohsawa, and Katsushi Ikeuchi. Separation of reflection
and transparency using epipolar plane image analysis. In *Proc. Asian Conf. on Computer Vision*,
pages 908–917, 2006.

[95] Michael Oren and Shree K. Nayar. Generalization of lambert's reflectance model. In *Proc. ACM
SIGGRAPH*, pages 239–246, 1994.

[96] Michael Oren and Shree K. Nayar. A theory of specular surface geometry. In *Proc. 5th Int. Conf.
Computer Vision*, pages 740–747, 1995.

[97] Pieter Peers and Philip Dutré. Wavelet environment matting. In *EGRW '03: Proceedings of the
14th Eurographics workshop on Rendering*, pages 157–166, Aire-la-Ville, Switzerland, Switzer-
land, 2003. Eurographics Association.

[98] A. Pentland, T. Darrell, M. Turk, and W. Huang. A simple, real-time range camera. In *Proc.
CVPR '89. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*,
pages 256–261, 1989.

[99] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical
Recipies in C*. Cambridge University Press, 1988.

[100] Stefan Rahmann and Nikos Canterakis. Reconstruction of specular surfaces using polarization.
In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern
Recognition*, volume 1, pages 149–155, 2001.

[101] Christian Reinsch. Smoothing by spline functions. *Numerische Mathematik*, 10:177–183, 1967.

[102] Arthur C. Sanderson, Lee E. Weiss, and Shree K. Nayar. Structured highlight inspection of
specular surfaces. *IEEE Trans. Pattern Anal. Machine Intell.*, 10(1):44–55, 1988.

[103] Silvio Savarese, Min Chen, and Pietro Perona. Local shape from mirror reflections. *Int. J. Computer Vision*, 64(1):31–67, 2005.

[104] Silvio Savarese and Pietro Perona. Local analysis for 3d reconstruction of specular surfaces - part ii. In *Proc. 7th European Conf. on Computer Vision*, pages 759–774, 2002.

[105] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1/2/3):7–42, 2001.

[106] Y. Y. Schechner, J. Shamir, and N. Kiryati. Polarization-based decorrelation of transparent layers: The inclination angle of an invisible surface. In *Proc. Seventh IEEE International Conference on Computer Vision The*, volume 2, pages 814–819, September 20–27, 1999.

[107] Yoav Y. Schechner, Shree K. Nayar, and Peter N. Belhumeur. A theory of multiplexed illumination. In *Proc. 9th Int. Conf. Computer Vision*, pages 808–815, 2003.

[108] Howard Schultz. Retrieving shape information from multiple images of a specular surface. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2):195–201, 1994.

[109] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition CVPR 2006*, volume 1, pages 519–528, 2006.

[110] Steven M. Seitz, Yashuyuki Matsushita, and Kiriakos N. Kutulakos. A theory of inverse light transport. In *Proc. 10th Int. Conf. Computer Vision*, pages 1440–1447, 2005.

[111] James Sharpe, Ulf Ahlgren, Paul Perry, Bill Hill, Allyson Ross, Jacob Hecksher-Sorensen, Richard Baldock, and Duncan Davidson. Optical projection tomography as a tool for 3d microscopy and gene expression studies. *Science*, 296:541–545, 2002.

[112] O.H. Shemdin. Measurement of short surface waves with stereophotography. In *Engineering in the Ocean Environment. Conference Proceedings*, pages 568–571, 1990.

[113] J.K. Solem, H. Aanaes, and A. Heyden. A variational analysis of shape from specularities using sparse data. In *IEEE Computer Society 3DPVT*, pages 26–33, 2004.

[114] O. Sorkine, D. Cohen-Or, Y. Lipman, M. Alexa, C. Rössl, and H.-P. Seidel. Laplacian surface editing. In *SGP '04: Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pages 175–184, New York, NY, USA, 2004. ACM.

[115] M. Subbarao. Parallel depth recovery by changing camera parameters. In *Proc. Second International Conference on Computer Vision*, pages 149–155, 1988.

[116] S. A. Sullivan. Experimental study of the absorption in distilled water, artificial sea water, and heavy water in the visible region of the spectrum. *Opt. Soc. Am. J.*, 53:962–968, 1963.

[117] Richard Szeliski, Shai Avidan, and P. Anandan. Layer extraction from multiple images containing reflections and transparency. In *Proc. Computer Vision and Pattern Recognition Conf.*, pages 246–253, 2000.

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

[119] Marco Tarini. Shape from distortion: 3d range scanning of mirroring objects. In *SIGGRAPH'02 Sketches and Applications*, page p.248, 2002.

[120] Marco Tarini, Hendrik P. A. Lensch, Michael Goesele, and Hans-Peter Seidel. 3D acquisition of mirroring objects. Technical Report MPI-I-2003-4-001, Max-Planck-Institut für Informatik, 2003.

[121] T. Treibitz, Y.Y. Schechner, and H. Singh. Flat refractive geometry. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition CVPR 2008*, pages 1–8, 2008.

[122] A. Treuille, A. Hertzmann, and S. M. Seitz. Example based stereo with general brdfs. In *European Conference on Computer Vision (ECCV)*, 2004.

[123] Borislav Trifonov, Derek Bradley, and Wolfgang Heidrich. Tomographic reconstruction of transparent objects. In *Proc. 17th Eurographics Symp. on Rendering*, pages 51–60, 2006.

[124] Bill Triggs, Philip F. McLauchlan, Richard I. Hartley, and Andrew W. Fitzgibbon. Bundle adjustment - a modern synthesis. In *Proc. ICCV '99: Proceedings of the International Workshop on Vision Algorithms*, pages 298–372. Springer-Verlag, 2000.

[125] P. Tu and P.R.S. Mendonca. Surface reconstruction via helmholtz reciprocity with a single image pair. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 1, pages I–541–I–547vol.1, 18-20 June 2003.

[126] Shinji Umeyama and Guy Godin. Separation of diffuse and specular components of surface reflection by use of polarization and statistical analysis of images. *IEEE Trans. on Pattern Anal. and Mach. Intell.*, 26(5):639–647, 2004.

[127] Y. Vasilyev, Y. Adato, T. Zickler, and O. Ben-Shahar. Dense specular shape from multiple specular flows. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition CVPR 2008*, pages 1–8, 2008.

[128] Hermann von Helmholtz. *Treatise on Physiological Optics*. Dover Publications, Inc., 1925.

[129] Jing Wang and K.J. Dana. Relief texture from specularities. *IEEE Trans. Pattern Anal. Machine Intell.*, 28(3):446–457, 2006.

[130] Y. Wexler, A. Fitzgibbon, and A. Zisserman. Image-based environment matting. In *Proc. Euro-Graphics Workshop on Rendering*, pages 289–299, 2002.

[131] R. J. Woodham, Y. Iwahori, and R. A. Barman. Photometric stereo: Lambertian reflectance and light sources with unknown direction and strength. Technical Report 91-18, University of British Columbia, Laboratory for Computational Intelligence, August 1991.

[132] Zhijian Wu and Guy A. Meadows. 2-d surface reconstruction of water waves. In *Engineering in the Ocean Environment. Conference Proceedings*, pages 416–421, 1990.

[133] Masaki Yamazaki, Sho Iwata, and Gang Xu. Dense 3d reconstruction of specular and transparent objects using stereo cameras and phase-shift method. In *ACCV 2007*, pages 570–579, 2007.

[134] Gang Zeng, Sylvain Paris, and Long Quan. Robust carving for non-lambertian objects. In *Proc. Int. Conf. on Pattern Recognition*, pages 119–122, 2004.

[135] Li Zhang, Noah Snavely, Brian Curless, and Steven M. Seitz. Spacetime faces: high resolution capture for modeling and animation. *ACM Trans. Graph.*, 23(3):548–558, 2004.

[136] Xin Zhang and Charles S. Cox. Measuring the two-dimensional structure of a wavy water surface optically: A surface gradient detector. *Experiments in Fluids*, 17:225–237, 1994.

[137] Jiang Yu Zheng and Akio Murata. Acquiring a complete 3d model from specular motion under the illumination of circular-shaped light sources. *IEEE Trans. Pattern Anal. Machine Intell.*, 22(8):913–920, 2000.

[138] L. Zhou, C. Kambhamettu, and D. B. Goldgof. Fluid structure and motion analysis from multi-spectrum 2D cloud image sequences. In *Proc. Computer Vision and Pattern Recognition Conf.*, volume 2, pages 744–751, 2000.

[139] Jiayuan Zhu and Yee-Hong Yang. Frequency-based environment matting. In *PG '04: Proceedings of the Computer Graphics and Applications, 12th Pacific Conference*, pages 402–410, Washington, DC, USA, 2004. IEEE Computer Society.

[140] T. Zickler, S.P. Mallick, D.J. Kriegman, and P.N. Belhumeur. Color subspaces as photometric invariants. In *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, volume 2, pages 2000 – 2010, 2006.

[141] Todd Zickler, Peter N. Belhumeur, and David J. Kriegman. Helmholz stereopsis: Expoiting reciprocity for surface reconstruction. In *Proc. 7th. European Conf. on Computer Vision*, pages 869–884, 2002.

[142] Todd Zickler, Peter N. Belhumeur, and David J. Kriegman. Toward a stratification of helmholtz stereopsis. In *Proc. Computer Vision and Pattern Recognition*, 2003.

[143] Douglas E. Zongker, David M. Werner, Brian Curless, and David H. Salesin. Environment matting and compositing. In *Proc. ACM SIGGRAPH*, pages 205–214, 1999.