Shape estimation of specular objects from multiview images

Visesh Chari

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Mathématiques, Informatique**

Arrêté ministérial : 7 Août 2006

Présentée par

Visesh Chari

Thèse dirigée par **Peter Sturm**
et codirigée par **Emmanuel Prados**

préparée au sein **INRIA Rhône Alpes, Laboratoire Jean Kuntzmann**
et de **Mathématiques, Sciences et Technologies de l'Information, Informatique**

Shape Estimation for Specular Surfaces

Contributions to Photogeometric and Multiview approaches.

Thèse soutenue publiquement le **20 November 2012**,
devant le jury composé de :

Prof. Edmond Boyer
Directeur de Recherche, INRIA Rhône Alpes, Président

Prof. Kyros Kutulakos
Professeur, University of Toronto, Rapporteur

Prof. Yoav Shechner
Professeur, Tehchnion, Rapporteur

Dr. Ivo Ihrke
Senior Recherche, MPI Informatik, Examinateur

Dr. Peter Sturm
Directeur de recherche, Inria Grenoble, Directeur de thèse

Dr. Emmanuel Prados
Chargé de recherche, Inria Grenoble, Co-Directeur de thèse

Contents

List of Figures

Abstract

The task of understanding, 3D reconstruction and analysis of the multiple view geometry related to transparent objects is one of the long standing challenging problems in computer vision. In this thesis, we look at novel approaches to analyze images of transparent surfaces to deduce their geometric and photometric properties.

At first, we analyze the multiview geometry of the simple case of planar refraction. We show how the image of a 3D line is a quartic curve in an image, and thus derive the first imaging model that accounts for planar refraction. We use this approach to then derive other properties that involve multiple cameras, like fundamental and homography matrices. Finally, we propose approaches to estimate the refractive surface parameters and camera poses, given images.

We then extend our approach to derive algorithms for recovering the geometry of multiple planar refractive surfaces from a single image. We propose a simple technique to compute the normal of such surfaces given in various scenarios, by equating our setup to an axial camera. We then show that the same model could be used to reconstruct reflective surfaces using a piecewise planar assumption. We show encouraging 3D reconstruction results, and analyse the accuracy of results obtained using this approach.

We then focus our attention on using both geometric and photometric cues for reconstructing transparent 3D surfaces. We show that in the presence of known illumination, we can recover the shape of such objects from single or multiple views. The cornerstone of our approach are the Fresnel equations, and we both derive and analyze their use for 3D reconstruction. Finally, we show our approach could be used to produce high quality reconstructions, and discuss other potential future applications.

Keywords: Computer Vision, Multiple View Geometry, Transparent Surfaces, Photometry, refraction, Reflection, Mirrors, Light-path Triangulation, Reconstruction, Fresnel Equations.

1 Introduction

Image understanding is a long standing fundamental goal of computer vision. The need to make sense of images has tremendous potential for use in various industries from robotics to surveillance. What makes the problem so hard is the complex interaction between the various physical elements of the scene (materials, light etc.) and the fundamental limitations of the camera (perspective projection, dynamic range etc.). One of the main components of image understanding is 3D reconstruction, and a lot of significant progress has been made in it in the last few decades.

Contents

We focus on 3D reconstruction of specular surfaces in this thesis. Surfaces with material properties that help them reflect or refract an incoming ray of light in a very narrow solid angle of directions are called specular surfaces (Figure 1.1). There are man-made and naturally occuring specular sufaces; they are ubiquitous and also widely used in industry. In this chapter, we first give a brief overview of 3D reconstruction with a bias towards specular surfaces (transparent and reflective). We argue that while a lot of progress has been achieved in "generic" 3D reconstruction, specular objects have been largely left out because their peculiar appearance does not sit well with most reconstruction algorithms. However, considering their ubiquitous nature (in fact, the camera lens itself is one example) any 3D reconstruction system will have to eventually deal with them, before being deployable on a large scale in the real world.

We look at physical properties of specular surfaces and devise algorithms that utilize these properties to reconstruct them from multiple images in the forthcoming chapters. While many existing works employ a variety of other sensors for this reconstruction task, we consciously choose to avoid them in our works. This is to facilitate a deeper understanding of the physics behind specular image formation, combined with an intention to have a wider set of scenarios where our algorithms and understanding might prove useful/insightful.

1.1 3D Reconstruction

3D reconstruction is the task of creating 3 dimensional models that faithfully comply with a set of 2D images of the object(s)/scene involved. Several approaches like structure-from-motion (sfm), stereo, photometric stereo are popular in the computer vision literature. While sfm and stereo are widely used for outdoor scene reconstruction, photometric stereo generally involves indoor scenes captured in controlled lighting environments. Current state-of-the-art 3D reconstruction algorithms perform remarkably well on both small and large scales. In fact, while reconstruction of entire cities

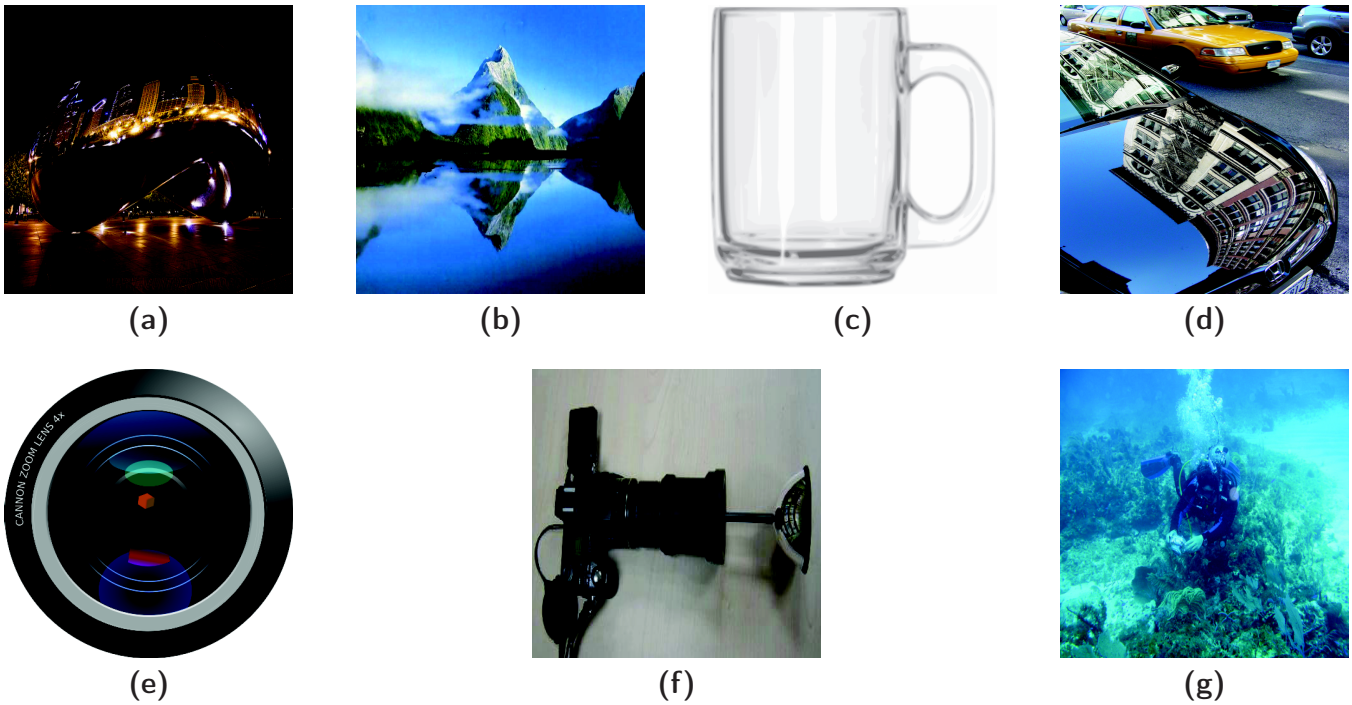


Figure 1.1: Different types of specular surfaces, and common applications. Man-made (a) and natural (b) specular surfaces. Commonly occuring (c) refractive and (d) reflective surfaces. Specular surfaces used in (e) cameras (f) catadioptrics (g) underwater photography. Figure (a) courtesy Andre Gunther. (http://www.agguntherphotography.com/great-lakes/chicago/millenium-park-bean.html). Figure (f) courtesy Maxime Lhuiller (http://wwwlasmea.univ-bpclermont.fr/Personnel/Maxime.Lhuillier/Omni3D.html).

has been recently attempted on the large scale side, reconstruction of mesostructures and extremely challenging objects like hair have also been attempted successfully.

One of the main obstacles that 3D reconstruction techniques have come across over the years has been the interaction of light and various materials, especially when it produces appearance varying results across images. In general, 3D reconstruction approaches rely on the Lambertian assumption of surface reconstruction. While in sfm and stereo algorithms, this results in the appearance constancy assumption (which is the basis of feature detection and matching algorithms), photometric stereo directly uses the Lambertian property to determine the relationship between pixel intensities and incident angle of the light rays from the light source. When the Lambertian surface assumption about material property fails, it introduces correspondence issues in sfm and stereo algorithms. Even in photometric stereo algorithms where correspondence is a non-issue, the failure of the Lambertian assumption causes reconstruction problems. To address this issue, various schemes have been proposed that take into account increasing generalizations of the Lambertian assumption.

In the case of specular objects, the Lambertian assumption is not at all valid. In fact, specularities have been argued to be characteristic of a specular object, and have been used for object recognition. As we show in the next chapter and as is known previously also, the multiple view geometry of specular surfaces too is very different from non-specular surfaces. This results in traditional sfm and stereo algorithms ignoring specular objects altogether, while traditional photometric stereo approaches have to undergo considerable modification before application to specular surfaces. Even 3D sensors like the ones used in Kinect perform poorly in the presence of specular surfaces.

One other challenge for specular objects is the modeling of their appearance. The appearance of specular objects not only depends on the material, but also on their shape, the incident illumination and the appearance of the scene in which they are placed. This makes it extremely difficult to use the appearance model of one specular object for the reconstruction of another. In fact, several different BRDF (Bidirectional Reflectance Distribution Function) models have been proposed to reconstruct different kinds of specular and near-specular surfaces; there are hardly any non-parametric approaches.

To summarize, two main obstacles prevent 3D reconstruction algorithms from being applied to specular surfaces and especially transparent surfaces.

- The Lambertian reflectance assumption is violated. This affects correspondence estimation in sfm/stereo approaches, and affects pixel intensity measurements in photometric stereo.

- Specular surface appearance depends on a variety of factors like object shape, incident illumination and surrounding environment. This results in the inability to produce a specific 3D reconstruction algorithm that applies to all specular surfaces.

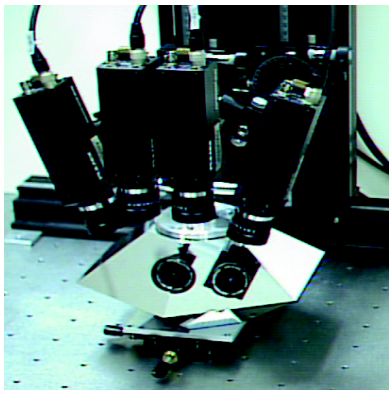
In the presence of these problems, it makes sense to look at algorithms that are specifically designed for specular surfaces. Our belief is that better image understanding could be obtained by investigating approaches that rely on a minimal amount of hardware, and thus we focus on a combination of geometry and photometry for specular surface reconstruction in this thesis. In fact, already some preivous works sharing our perspective of image understanding can be seen in [117] where the authors propose photometric invariants useful for detection and recognition of transparent surfaces that are based on geometric properties of smooth surfaces. In the next section, we give a brief history of the various approaches employed for reconstruction of specular surfaces to set the context for our works. Additionally, we list and briefly outline the various contributions in this thesis.

1.2 History with a Specularity Bias

The history of specular object reconstruction is divided in two parts. On one hand, the reconstruction (calibration) of catadioptric cameras has been important for image acquisitions of a specific kind used in robotics, panorama creation etc. The reconstruction of transparent specular surfaces was first initiated in photogrammetry for use in underwater photography [83]. More recently, in computer vision transparent object reconstruction has been attempted using physical properties of transparent surfaces like their specular property and polarization.

1.2.1 The Case of Mirrors

Figure 1.2 shows some catadioptric systems, the first class of mirrors that were used in computer vision. Their primary usage was to enhance the field of view of conventional cameras, to obtain wide fields of view (FOV) [8]. Using a mirror with a camera required ”calibration” of the mirror, which in turn meant knowing the shape parameters and position of the mirror w.r.t the camera. Thus calibration implicitly involved 3D reconstruction of the surface, although in a restricted sense since the shape model (hyperbolic, parabolic etc.) was already known in these cases. The primary purposes of using catadioptric systems was for 1) Panoramic imaging [136] and 2) SLAM applications [69].



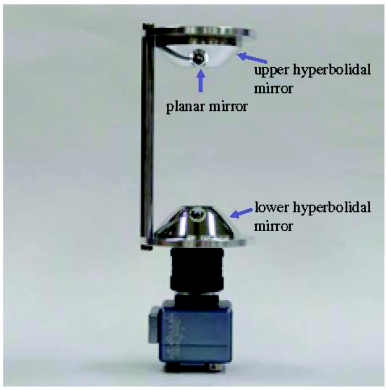


Figure 1.2: Catadioptric systems are widely used for panoramic image construction and visual SLAM. Two designs of catadioptric systems are presented in (a) [136] and (b) [69]. (c) A typical image captured by a catadioptric camera (rectified image shown on the top) [76]

Generic specular surface reconstruction was first proposed in [14], where the authors considered reconstructing the surface by observing highlights in stereo images. In general, the first set of reconstruction algorithms noted that appearance of specularities in different images could contain clues to local shape information [13]. These works were predated by the seminal work of [70], which outlined the different photometric invariants related to reflective surfaces. Since specularities are produced by illumination reflected off the specular surface, photometric stereo methods [68] were also extended to specular surface reconstruction. Finally, since it was realized that motion of specularities contain information about local shape [13], a natural extension was to use not just specularities but motion of all reflected points on the surface of a mirror for its reconstruction. This lead to a "shape from distortion motion" area of approaches [1; 115; 118].

Catadioptrics One of the earliest uses of spherical mirrors was demonstrated in [101], where the authors used two spheres of known radii, placed at known positions w.r.t the camera in order to compute the depth of the world using a single image. This was one of the first uses of a non-single viewpoint camera for depth estimation, made easy because of the simplicity of a spherical setup and manually measured position estimates. Subsequently, the semilar work of single viewpoint catadioptric cameras [8] generalized the kind of mirrors that could be used in a catadioptric setup while ensuring that the entire system had a single effective viewpoint. Since using such mirrors with cam-

eras requires knowledge of their shape and position w.r.t the camera, several ”calibration” algorithms were proposed to acquire these parameters [38; 39]. These calibration methods determined the shape parameters, given known shape models and constituted one of the first reconstruction algorithms for mirrors. The single-viewpoint assumption proved useful because it allowed extending traditional multiple view geometry [55] to the case of catadioptric setups.

This approach was later extended to include non-central viewpoints *i.e.* camera-mirror setups that were not single viewpoint systems. The advantages of such systems included increased control over FOVs (Field of View) and general better designing flexibility [134]. Calibrating such cameras could now involve specific algorithms that took into account shape models [41] or could take the form of general calibration of a system of rays [131]. While the former approaches performed reconstruction, the latter just reconstructed rays and the eventual shape of the mirror could be obtained by triangulation of captured ray directions and incident ray directions in 3D. A recent survey and taxonomy of non-central camera calibration can be found in [123]. While many of the previous approaches could be restructured to get 3D reconstruction of arbitrarily-shaped mirrors [82; 112; 131], reducing the number of 2D-3D correspondences gives an unstable solution [15]. Since it might not be practically feasible to get many images using for the same mirror-camera configuration in many general settings, it might be worthwhile looking at how different mirrors might be reconstructed using piecewise modeling. We investigate this idea in the fourth chapter.

Specular Flow Methods When a moving camera observes a specular surface, or when a mirror-camera setup moves in the world, the appearance of the specular surface captured by the camera changes over time [154]. The optical flow obtained from the multiple images captured in such a setup is called *specular flow* [115]. A part of this problem has been known from a long time as Alhazen's Billiard's problem [127]. The deviation of *specular flow* approaches from previous attempts at this area was the relaxation of the knowledge of the illumination source [13; 14; 154]. Thus it directly retrieved camera motion and mirror shape by just observing reflections off the surface. While progress has been made in this field, specular shape from flow algorithms still provide results that are below the state-of-the-art when compared to other approaches that assume known illumination sources (or known 2d-3d correspondences, because they serve the same functionality in the case of mirrors) [16; 17; 82]. Some notable approaches in this direction are [1; 118; 143]. In fact, it is useful to note that some of these algorithms are related to non single viewpoint cameras [143]. It

might also be worthwhile investigating whether the difficulty in specular flow based approaches is in any way connected to the instability of pose estimation of a checkerboard pattern in the presence of a generic mirror, when a direct view of the checkerboard pattern is not available [15]. Finally, while specular flow approaches exploit the motion information produced in images and their relationship to mirror shape and world motion, we restrict the focus in this thesis to static instances of specular surface reconstruction approaches. The underlying motivation for us is two-fold

- We believe that reconstruction of surfaces from static setups has not been given much attention. In fact, many of our approaches just require a single view (not to be confused with single image) and in some cases, we actually are able to obtain more information about the object shape than traditional 3D reconstruction scenarios (with opaque objects).

- We believe that the understanding obtained from solving and thinking about such setups gives us more insights into the general cases of shape from specular flow. More importantly, the relationship between specular flow and non-single viewpoint cameras is well worth exploring, although in the current thesis this direction is not explored. We however explore related areas that might give insights into the nature of this connection.

1.2.2 The Case of Transparency

Transparent objects are one of the few categories in which reconstruction methods are still in their nascent stage in terms of accuracy, ease of use etc. Traditional shape estimation techniques make assumptions that are rendered invalid when the image formation process of transparent objects is considered. A recent survey [67] indicates the various approaches used to acquire the shape of transparent objects. Figure 1.4 lists the various categories which are considered in this survey. In this section, we give a brief description of some of these approaches, as well as others used in photogrammetry [83].

Active Methods One of the most accurate object shape acquisition techniques in computer vision is structured light projection. Active methods like these use an external stimulus (like projector light) to modify the image acquired by the camera, in a way that makes it easier to compute the shape of the object of interest. While it is a highly accurate technique in the case of opaque objects, there

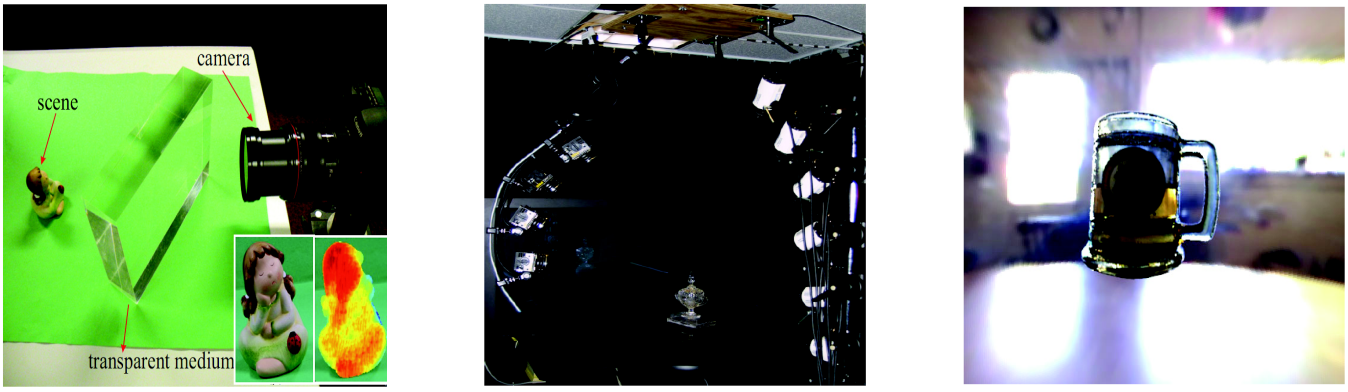


Figure 1.3: (Left) Transparent surfaces with known 3D structure may be used for 3D reconstruction of other objects [26] (object reconstructed depthmap in inset). (Middle), (Right) Acquisition of transparent surfaces is important for realistic rendering [85]

object type	surface / volume type	class	image formation
opaque	surface, rough	1	diffuse or near diffuse reflectance
	surface, glossy	2	mixed diffuse and specular reflectance
translucent	surface, smooth	3	ideal or near ideal specular reflectance
	surface, sub-surface scattering	4	multiple scattering underneath surface
transparent	surface, smooth	5	ideal or near ideal specular refraction
	volume, emission / absorption	6	integration along viewing ray
	volume, single scattering	7	integration along viewing ray
	volume, multiple scattering	8	full global light transport without occluders
inhomogeneous	mixed scenes, containing many / all of the above	9	full global light transport

Figure 1.4: The different classes of transparent surfaces as described in [67]

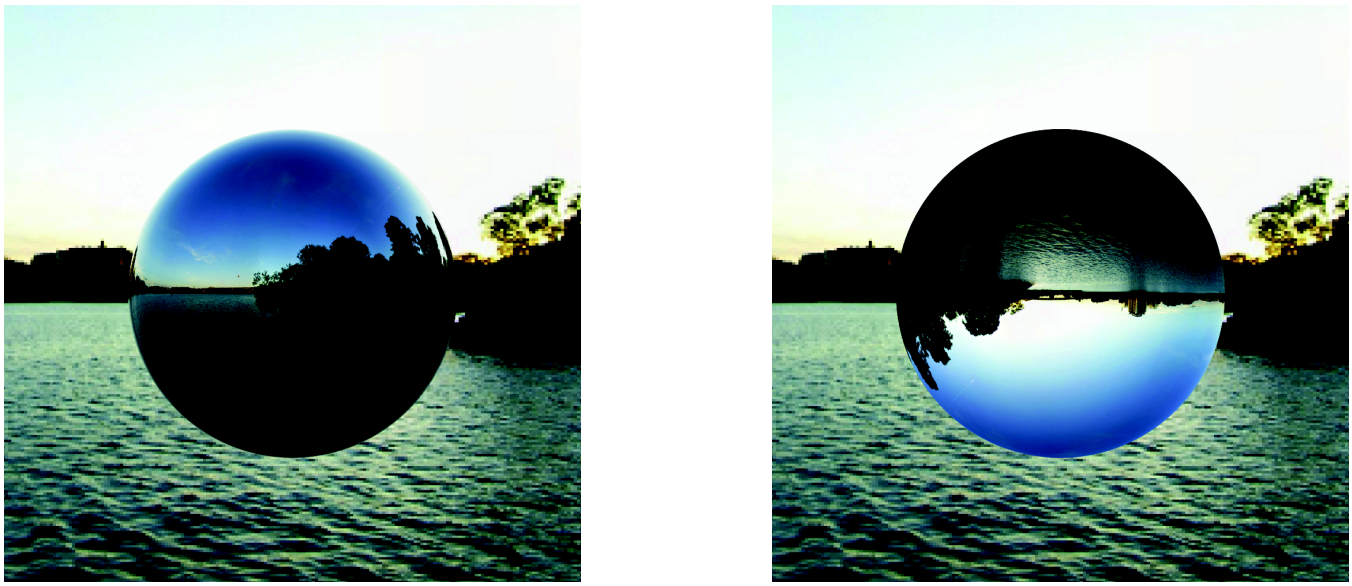
have only been recent attempts at active “scanning” of transparent surfaces. This is because of the presence of optical phenomenon like sub-surface scatterring, interreflections, scattering (in the case of participating media) etc. Recent approaches like [91] have extended active techniques to transparent surfaces by modeling their surface properties so as to isolate their reflective capabilites from other optical phenomenon, and use it for reconstruction.

Polarization Polarization is a physical property of surfaces that has been used for reconstruction purposes. When an unpolarized beam of light is incident on a reflective/refractive surface (not necessarily a mirror), its reflected and refracted components are partially polarized, if the object is dielectric. Since the angle of polarization is directly related to the ratio of incident and reflected/refracted light through Fresnel equations [36], determining the polarization angle gives useful information about the shape of the object. In fact, earlier approaches have used a camera with a linear polarizer to determine this angle, and subsequently used it for reconstruction [65]. Recently, many papers have used this approach to reconstruct transparent surfaces [88; 89]. This extension normally requires a relatively sophisticated setup, and both forward and inverse (predicting the shape and comparing it with the image obtained in order to refine the shape) methods have been proposed in this area [7; 116]. Other approaches have also been reported in the literature that have used polarization to reconstruct metallic mirror-like surfaces [90].

Shape from X While several of the approaches mentioned earlier (like specular flow) could come under the category of shape from X, we choose to put methods that are not directly relevant to this thesis, but still related, in this category. Figure 1.3 (middle) gives one such setup [85]. The authors propose an entire system which segments and starts with a coarse reconstruction of a transparent surface irrespective of its transluscency (several categories in Figure 1.4). This is then refined using the several images obtained with various light sources, as shown in the setup (Figure 1.3, middle image). Some other approaches ”acquire” the shape of transluscent surfaces, like the ones depicted in Figure 1.7. In [52], the authors recover the structure of inhomogeneous media like milk by using compressed sensing based algorithms to analyze the images obtained after structured light is projected onto a setup consisting of the object of interest immersed in media like water.

Photometric Methods Traditional photometric methods employ a static viewpoint (relative position of camera and object is fixed), and move the light source to generate varying appearance of the object [145]. While several assumptions about photometric stereo have been relaxed subsequently, the first application to specular surfaces was done by [68]. The approach used in this method considered a distance light source and a single-view approach in which multiple images under changing illumination were obtained. Surface orientations were then recovered using this data. Further extensions to this approach were done either using exemplars [61], color [84] using Helmholtz theories [155] etc. The general idea in photometric stereo approaches has been to consider specular surfaces as surfaces with unique BRDF's, which means that moving away from the Lambertian assumption is necessary [84]. While the previous approaches extend the photometric approach to specular surfaces, they are restricted to mirror like BRDF's. In the presence of transparent surfaces, additional optical phenomena like interreflections, scattering, sub-surface scattering, chromatic aberration are involved [91], because of which the photometric data obtained is noisy (with bias). Unfortunately, the general approach of photometric stereo does not provide the scope for removing such errors. In such a case, several approaches like polarization based approaches [7; 88], or robust specularity estimation based approaches [148] or physical model based approaches that can detect and remove such errors [91] have to be used. Note that while these approaches use light sources in order to estimate the shape, some like [91] do not use a BRDF based approach to extract surface normals based on the intensity of light image by the camera at a point, and so they can only be related to photometric stereo methods by a distance.

Light-Path Approaches A light-path is defined as the path taken by a ray of light from the point of emergence till the time when it reaches the camera [73]. When the scene consists of specular surfaces, this path is piecewise-linear 3D, although it may not be planar (this distinction becomes important in the fourth chapter). Thus specular surface reconstruction could be alternately defined as reconstruction of *all* the light-paths that pass through the scene and are captured by the camera. While not explicitly stating so, several approaches have attempted this in the past, or have light-path reconstruction as a reasonably simple extension [3; 17; 82; 112; 131]. One of the major problems with this type of approach is that it is inherently local in nature, in the sense that so far, most approaches in the literature focus on reconstructing individual light-paths separately and a global algorithm for reconstructing all light-paths simultaneously with the shape of the object(s) is missing [73]. The



(a) (b)

Figure 1.5: Examples considered in this thesis. Depiction of a (a) reflective sphere, and a (b) refractive sphere. Note that the refractive sphere represents two refractions. Our goal is to reconstruct the shape of these specular objects (without assuming a model like a sphere), using information from the surrounding environment and images.

authors in [73] also impose a limit on the number of bounces (number of piecewise linear segments) a light-path can have if its reconstruction should be possible. However, we believe that light-path reconstruction is closely related to non-single viewpoint geometry, and reasoning about both can lead to solutions to cases where there are more than two bounces inolved per light-path. In fact, we extend this limit in some cases in the third and fourth chapter.

Summary To summarize, there is an extensive literature in specular object reconstruction. While many approaches have focused on extending traditional reconstruction methods to the case of specular surfaces, the most successful methods currently focus on using algorithms tailor made to the characteristics of specular surfaces in order to produce reasonable reconstructions [35]. In the next section, we outline the specular reconstruction problem, and give a brief overview of our line of reasoning, while presenting our contributions.

1.3 The case of specularity

In this section, we define the quintessential specular surface reconstruction problem considered in this thesis. Figure 1.6 depicts the scenario that we consider, partially, while 1.5 gives two examples of the kind of images we deal with. A known / unknown 3D point p is reflected/refracted off several surfaces, at points v_i before being finally imaged as point **q** at the camera **c**. This light-path may or may not be accompanied by a radiance measurement which in the case of the transparent surface, would involve its refractive index relative to the outside medium, apart from the angle of incidence of the incoming light ray and the shape (normal) of the object itself. Finally, we can consider a single light-path in isolation or multiple light-paths together. Our objectives are many-fold.

- To derive basic understanding of the interplay between transparency and light-paths. To further relate it to the image formation process and the multiple view geometry involved.
- To use this understanding for the reconstruction of transparent surfaces from multiple images. To extend this understanding to general specular surfaces (i.e. mirrors too) and further relate them to existing understanding on the specular image formation process.
- To use additional input (like radiance) readily available without additional equipment (polarizers) in order to investigate the extent to which light-path triangulation [73] may be solved in several cases.
- To identify new problems and scopes of extension of works established in this thesis.

1.3.1 Reconstruction vs Acquisition

Figure 1.7 shows several systems aimed at "acquiring" properties of transparent surfaces. In many cases, like the work of [85], the acquisition involves reconstructing the object involved, while in many others like [97] only the optical properties of the object are desired. Most methods of acquisition are for purposes of using the material in graphics applications [64; 85; 97]. While acquiring properties accurately is a hard process, approaches with graphics applications in mind tend to ignore the image understanding perspective and employ several additional equipment like a fluorescent dye in the case of [64]. However, our argument is that reconstruction based approaches [73; 91] etc. are

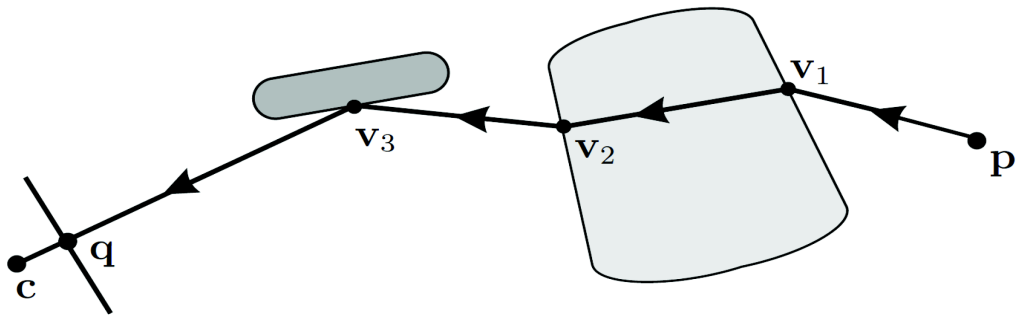


Figure 1.6: Depiction of the specular surface reconstruction problem from [73]. A known/unknown 3D point p undergoes 1 or more reflections/refractions before being imaged by a camera **c**.

"minimalist" in their usage of additional equipment because of the need for application to outdoor scenarios where a robot with a camera or a mobile vision platform might not have access to such equipment. We believe that our approaches help us better serve these goals because of the following reasons.

- We model the physical nature of reflection/refraction in order to obtain the 3D reconstruction of scenes. This is exactly the extension of general 3D reconstruction algorithms (since appearance of specular surfaces is reflective/refractive in nature) to the case of specularity.

- 3D reconstruction of objects considered in this thesis results from an understanding of the interactions between light and object material, and their resulting effects on image formation. Thus, data acquired from our setups, for example, could be used to understanding the nature of interreflections within transparent surfaces.

- As opposed to acquisition techniques, we restrict our study to specular surfaces. This allows us to focus on a particular set of optical phenomenon, while ignoring other complex ones like scattering. This isolation enables better understanding.

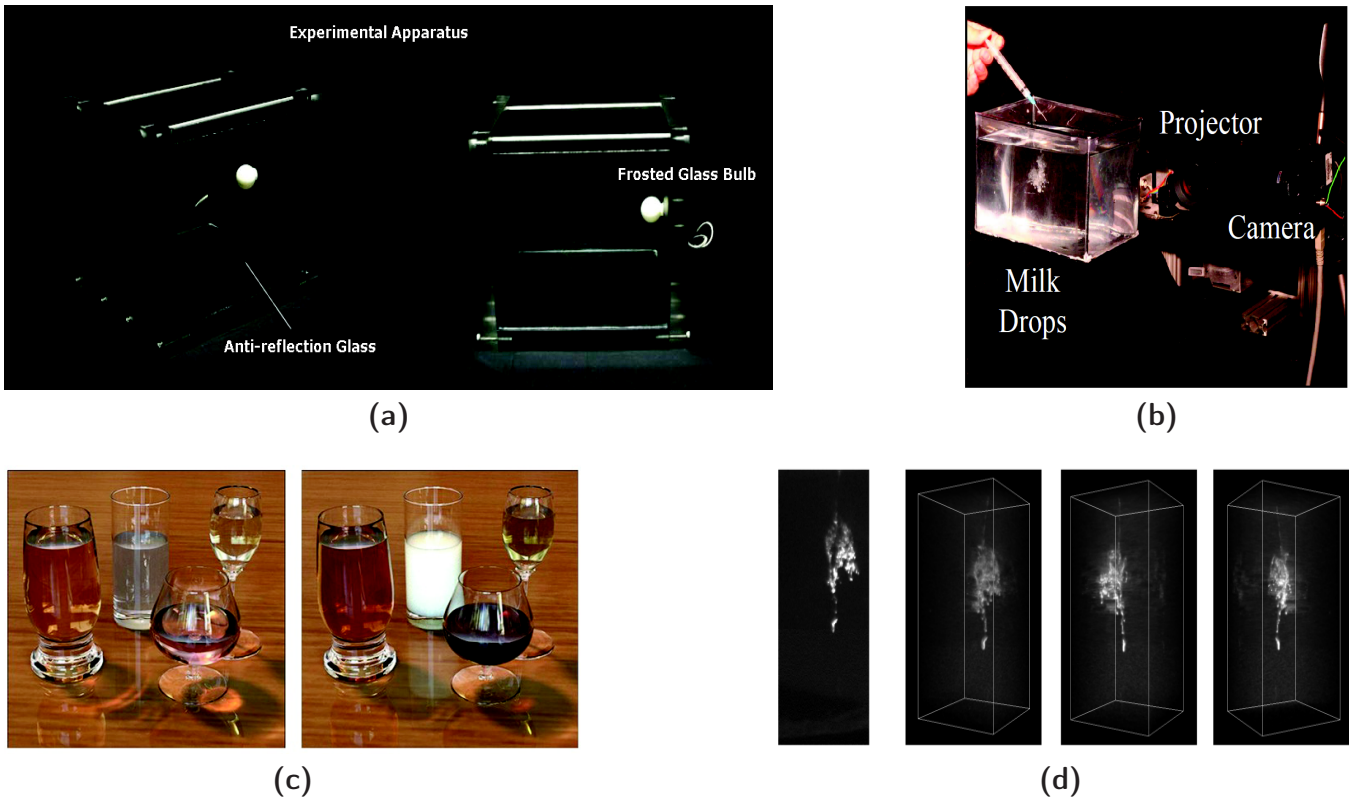


Figure 1.7: Acquisition of transparent surfaces [52; 97](Top row) Setups to acquire participating media (wine) and inhomogeneous media (milk in water). (c) Rendering using acquired material properties. (d) One of the acquired images, and reconstruction shown in 3 views.

1.3.2 Physical Modeling vs Light-Path Modeling

Figure 1.8 shows results from various reconstruction approaches in the literature [73; 88; 91]. These reconstruction approaches can also be divided into two parts, based on the ideas and models they use. Physical modeling based approaches [68; 88] try to extend non-triangulation based methods (in many cases, variational approaches like [128]) to the case of specular surfaces. While such approaches also produce encouraging results, we stick to the light-path triangulation perspective for a few reasons.

- **Acquisition Ease** As we show in the later chapters, our acquisition setups are generally much simpler and data easier to acquire (without additional equipment like polarizers and diffusers) than current physical modeling based approaches.

- **Usage of Inputs** While being easier on the image acquisition side, we also end up utilizing all properties of specular surfaces (like polarization, radiance ratios etc.) that most of the physical modeling approaches use. In fact, our models might also be categorized as physical models within a light-path framework.

1.4 Our Contributions

In the previous sections, we have outlined our reasoning for choosing this perspective for studying specular object reconstruction. In this thesis, we make the following contributions to this line of study.

Chapter 2 In the first part of this thesis, we analyze the multiple view geometry of a refractive planar surface. We consider the case when one or multiple cameras in a medium (for example air) are looking at a scene in another medium (for example water), with the interface between the two media being flat. The case of underwater photography fits this description. Since a perspective projection model no longer fits this scenario, at first we derive the forward projection model and related camera matrix. We show that 3D lines in a scene map to quartic curves in the image in such a scene. An interesting observation about this scenario is that there is a unique image curve for every 3D line in the world, assuming a homogenous refractive index. We then derive multiple view quantities like the fundamental and homography matrices related to this scene and count arguments for tasks like pose estimation of cameras from multiple images. We also show that when the camera is in a heavier

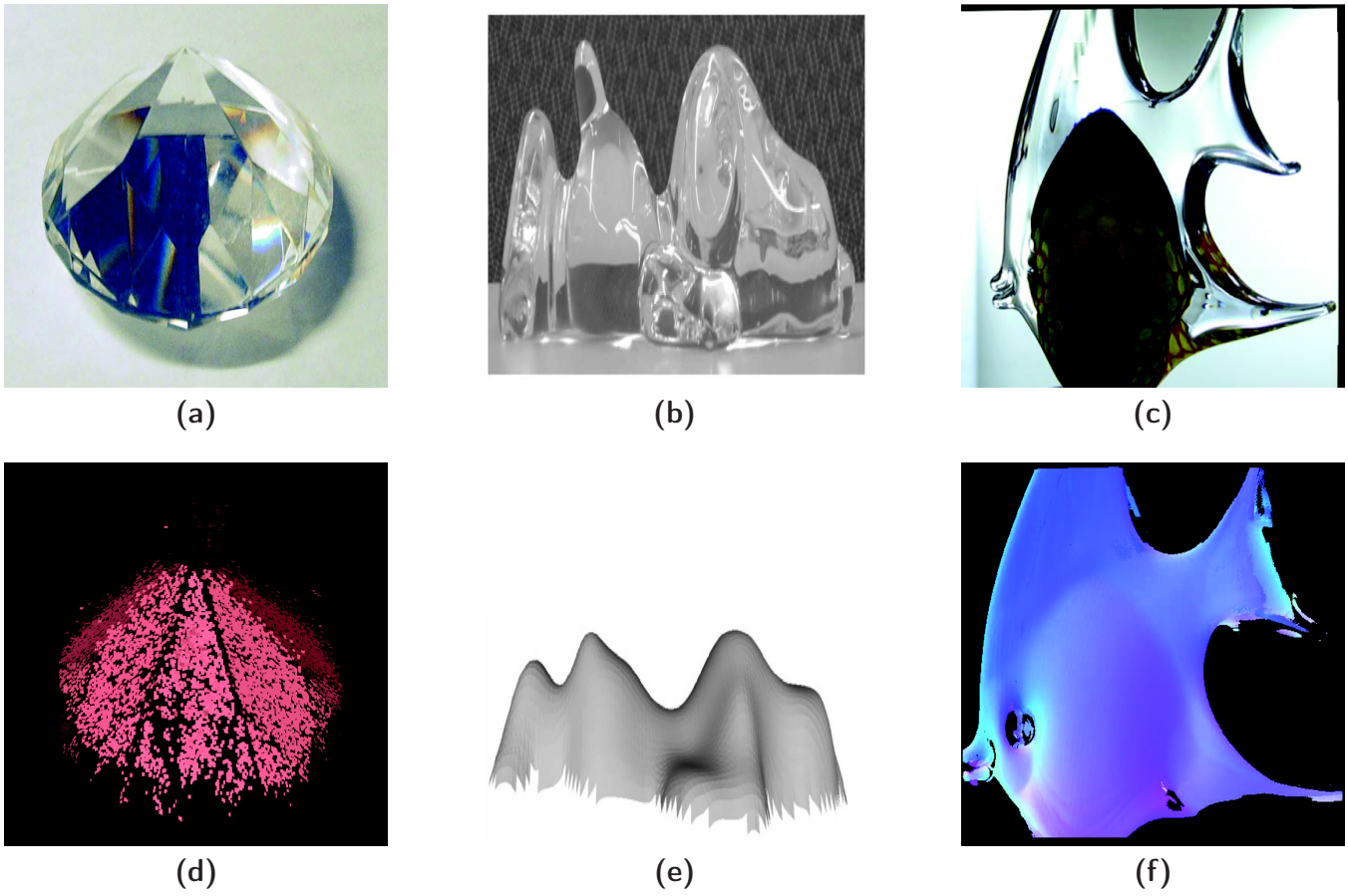


Figure 1.8: Reconstruction of transparent surfaces. (Top row) Setups to reconstruct transparent objects ([73; 88; 91] in order). (Bottom row). 3D point cloud, mesh and normal map of the reconstructed objects produced using the various methods.

medium, the horizon maps to a conic ("Snell's window") which can be decomposed to obtain the parameters of the separating interface.

Chapter 3 Transparent specular objects also modify and re-distribute the light energy incident on their surfaces. The governing model of this re-distribution is represented by Fresnel's equations. Using this as our cornerstone, we describe a method that combines both geometric and photometric information to do reconstruction of arbitrarily shaped transparent specular surfaces. We show that our approach leads to a very uncomplicated acquisition process, while keeping our approach fairly simple. First, we analyze several minimal cases for shape reconstruction, and derive novel constraints for reconstruction that combine geometric and Fresnel theories about transparent surfaces. Secondly, we illustrate the complementary nature of these cues which helps us gain additional information about the object, which is otherwise impossible to obtain. Finally, we discuss practical aspects of our reconstruction algorithm while presenting reconstruction results on challenging datasets. Our results show that high quality reconstructions can be achieved in challenging scenarios.

Chapter 4 We then extend our understanding of planar refractive geometry to derive algorithms for recovering the geometry of multiple planar refractive surfaces from a single image. A typical example of such a scenario would be looking through a fishtank, or underwater photography. We propose a simple technique to compute the normal of such surfaces given in various scenarios, by equating our setup to an axial camera. This allows us to fit RANSAC based approaches like the 8-point algorithm for fundamental matrix computation to our case, in a manner similar to axial distortion estimation techniques in the computer vision literature. Interestingly, this approach naturally extends to *planar light-paths* within axial systems. To this extent we then show that the same model could be used to reconstruct reflective surfaces consisting of multiple parallel planes or axially aligned spheres. We show encouraging 3D reconstruction results, and analyze the accuracy of results obtained using this approach.

Conclusion In conclusion, we analyze several different geometric and photometric properties of specular transparent surfaces and explore two ends of a reconstruction spectrum: shape specific reconstruction and shape independent reconstruction. In both cases, we present several novel theoretical and algorithmic contributions with encouraging results. We show how the perspective of light-

path triangulation could be extended for better image understanding of specular surfaces. Potential applications include robotics [69] and graphics [85], applications like specular object detection and recognition [105; 117].

2

Multi-View Geometry of the Refractive Plane

2.1 Introduction

One of the ways in which we can group and study light-paths is to consider grouping light-paths that are the projections of a 3D point or line on a camera, after undergoing several reflections, refractions or both. In this chapter, we consider the simplest case of such a scenario where the imaging characteristics of a camera in one medium looking at a scene in a different medium with a planar, transparent interface between the two is studied.

The multi-view geometry resulting from opaque scenes is now well understood, for the case of perspective projection. To some extent, even the insertion of reflective elements has been studied in the area of catadioptrics [130; 132]. The phenomenon of refraction, however, has largely been left un-addressed in the vision community.

Contents

The introduction of refractive elements into a scene changes the multi-view geometry that results from the imaging phenomenon. Until recently, the study of this change had been restricted to multimedia photogrammetry [83] and oceanic engineering [102], where the major perspectives were to either neglect the refraction [50; 102], view refraction as an error or aberration to perspective imaging [107; 126], or to look at its correction as an iterative optimization problem [83]. In computer vision, some of the first attempts have come in the recent past [9; 73; 74; 92; 141]. Kutulakos et al. [73] investigate the geometry of *light-path triangulation*, which aims to find conditions and algorithms where reconstruction of individual tracks of light is possible. The work of Morris and Kutulakos [92] looks at refractive stereo. The main idea here is to estimate the normal of the refractive surface, given 2D-3D correspondences, irrespective of the refractive index. They argue that under the assumption that light is refracted only once, two views are sufficient to reconstruct an arbitrary

refractive surface. The work of Nayar et al. [9] concentrates on the estimation of the shape of the object by optimizing a function that minimizes the difference between observed and predicted images, based on a suitable parametric model of the object shape, as well as assuming an affine setting. The works of Singh et al. [141] and Lavest et al. [74] concentrate on the internal calibration of cameras underwater, when a planar refractive surface separates the two media (air and water).

However, to the best of our knowledge, the epipolar geometry resulting from refraction has not been analyzed till now. In the presence of refraction, this geometry not only encodes the relative position of the cameras, but also information about the relative refractive index between the two media in question, and the geometry of the surface separating the concerned media. In this paper, we concentrate on the multi-view relationships induced by a refractive planar surface. We choose to deviate from the generic scene assumption of Kutulakos *et al* [73], which enables us to completely describe the epipolar geometry in the presence of refraction. In this respect, our work is closer to Barreto & Sturm [130] and Glaeser & Schröcker [44]. In [44], the authors define the equation that governs the projection of a 3D point onto an image, while being refracted across a planar surface. We follow an approach that is along the lines of [130] in deriving the various multi-view relationships in the presence of refractions by a planar surface whereas did so for [130] did so for reflection off quadrics (central catadioptric cameras).

In the following sections, we derive the various multi-view relationships between two views of a scene, when a planar transparent surface separates the scene from the camera. The two media are assumed to have different refractive indices. First, we define the projection matrix for such cameras for lines, where we show that a 3D line is mapped to a quartic curve. Next we derive the *fundamental* matrix between two views, which turns out to be a function of the camera poses and the plane normal and refractive index. We then derive an expression for the homography between a scene plane and its image on the camera. Finally, we give algorithms for obtaining the relative pose between two cameras in specialized cases, by deriving the expression of what is commonly referred to as *Snell's Window*.

Background & Notation Consider the scenario where the scene contains a planar transparent refractive surface, with cameras on one side of the plane and the actual scene, composed of opaque objects, on the other side. Let the relative refractive index between the two media on either side of the plane be denoted by λ ($= \frac{n_1}{n_2}$), where n_1 is in the medium with cameras). Let us remember that

the incident and refracted angles are related by

$$\sin(\theta_2) = \lambda \sin(\theta_1) \tag{2.1}$$

where θ_1 is the incident angle, and θ_2 is the refracted angle.

Images of objects on the opposite side of the cameras are formed after refraction by the surface in consideration. We wish to deduce external calibration information about the cameras and the geometry of the refracting plane from such images alone. Further along this chapter, we make certain assumptions about the geometry of the scene, refractive index etc. to simplify matters. Such assumptions will be detailed when necessary.

In the rest of the chapter, we use the following coordinate system. Let the 3D refractive plane, denoted by $\pi = \begin{pmatrix} \mathbf{v}^\top & d \end{pmatrix}^\top$ take the values $\begin{pmatrix} 0 & 0 & 1 & 0 \end{pmatrix}^\top$, which aligns it with the X-Y plane of the world coordinate system. Let two cameras P_1 and P_2 be situated in the following manner

$$\mathsf{P}_1 = \mathsf{R}_1 \begin{pmatrix} \mathsf{I} & -\mathbf{t}_1 \end{pmatrix} \tag{2.2}$$

$$\mathsf{P}_2 = \mathsf{R}_2 \begin{pmatrix} \mathsf{I} & -\mathbf{t}_2 \end{pmatrix} \tag{2.3}$$

$$\mathbf{t}_1 = \begin{pmatrix} 0 & 0 & -1 \end{pmatrix} \tag{2.4}$$

where $\mathsf{R}_1, \mathsf{R}_2$ are rotation matrices. Here, the position $\mathbf{t}_1$ of the first camera is fixed without loss of generality. The position $\mathbf{t}_2$ of the other camera, is not restricted.

Now, the only undecided component about the world coordinate system is the location of the X- and Y-axes, which can vary upto rotation about the Z-axis. Although this can easily be fixed by assuming the first camera's coordinate axes to be aligned with that of the world, we ignore this unnecessary assumption for now. Finally, for the purposes of this chapter, the internal parameters of all cameras are assumed to be $\mathsf{I}_{3\times3}$,i.e., the cameras are calibrated.

Now, let us define notations for points. We denote an image point by the variable $\mathbf{x}$, which is a 3×1 homogeneous vector defined as $\mathbf{x} = \begin{pmatrix} x & y & 1 \end{pmatrix}^\top$. The direction vector for an image point is a 4×1 homogeneous coordinate and is represented by $\mathbf{q} = \begin{pmatrix} \tilde{\mathbf{q}}_{3\times1} & 0 \end{pmatrix}^\top = \begin{pmatrix} (\mathsf{R}^\top \tilde{\mathbf{x}})^\top & 0 \end{pmatrix}^\top$, and its j^{th} coordinate for the i^{th} camera is represented by $q_{i,j}$, or by q_j when there is only one camera. The vector $\tilde{\mathbf{x}}$ represents the unit vector corresponding to image point $\mathbf{x}$. Thus $\|\tilde{\mathbf{q}}\| = 1$. A vector formed by

collecting elements from another is represented in braces. For example, $\mathbf{L}_{1,(6,1,2)}$ represents a vector formed by elements 6, 1 and 2 in $\mathbf{L}_1$. The Plücker coordinates of a 3D line is denoted by $\mathbf{L}$. Matrices are represented using the font M, and vectors are represented as $\mathbf{v}$. $\widehat{\mathbf{M}}$ denotes *lifted* coordinates ($\widehat{\mathbf{q}} = \begin{pmatrix} q_1^2 & q_1q_2 & q_2^2 & q_1q_3 & q_2q_3 & q_3^2 \end{pmatrix}^\top$ [63; 130]). Finally, if two vectors are related by a linear transformation ($\mathbf{q} = \mathsf{R}^\top\mathbf{x}$), their lifted coordinates are related by [63; 130]

$$\widehat{\mathbf{q}} = \mathsf{D}^{-1}\mathsf{S}(\mathsf{R} \otimes \mathsf{R})\mathsf{S}^\top\widehat{\mathbf{x}} \tag{2.5}$$

Henceforth, all derivations are done with $\mathbf{q}$, *i.e.* the back-projection direction for image point $\mathbf{x}$, while noting that the corresponding equations for $\mathbf{x}$ can be obtained (if desired) by using Equation 2.5.

2.2 The Camera Projection Matrix

The first step in deriving the epipolar geometry needed to describe multi-view relationships is the camera matrix. We proceed to derive the camera matrix by first defining the back-projection of a point, and then focusing on the condition that it intersects a 3D line.

2.2.1 Back-projecting a Point

It has been previously shown that the image of a 3D point after refraction from a plane is the solution of a quartic equation that lies within a specified interval [44]. We proceed to show that the image of a line after refraction is a quartic curve, and from this projection we derive a description of the camera matrix for refraction.

Consider a 3D point with inhomogeneous coordinates $\mathbf{X}$. An image point $\mathbf{x}$ with back-projection direction $\mathbf{q}$ (both in homogeneous coordinates) represents an image corresponding to $\mathbf{X}$, if the line from the camera center in the direction of $\mathbf{q}$ passes through $\mathbf{X}$ after refraction. After developing the refraction equation 2.1 for vectors, we can get the point $\mathbf{Q}$ on the refraction plane π, and the direction

of the refracted ray $\mathbf{R}$ as

$$\mathbf{Q} = \begin{pmatrix} t_3\tilde{\mathbf{q}} - q_3\mathbf{t} \\ -q_3 \end{pmatrix}, \qquad \mathbf{R} = \begin{pmatrix} q_1\lambda \\ q_2\lambda \\ \sqrt{(1-\lambda^2)+\lambda^2 q_3^2} \\ 0 \end{pmatrix} \tag{2.6}$$

where $Q_3 = 0$. The corresponding Plücker coordinates of the refracted ray can be written as

$$\mathbf{L} = \begin{pmatrix} Q_4\mathbf{R}_{1\ldots3} - R_4\mathbf{Q}_{1\ldots3} \\ \mathbf{Q}_{1\ldots3} \times \mathbf{R}_{1\ldots3} \end{pmatrix} = \begin{pmatrix} -q_1 q_3 \lambda \\ -q_2 q_3 \lambda \\ -q_3\sqrt{(1-\lambda^2)+\lambda^2 q_3^2} \\ (q_2 t_3 - q_3 t_2)\sqrt{(1-\lambda^2)+\lambda^2 q_3^2} \\ (q_3 t_1 - t_3 q_1)\sqrt{(1-\lambda^2)+\lambda^2 q_3^2} \\ (q_1 t_2 - q_2 t_1) q_3 \lambda \end{pmatrix} \tag{2.7}$$

Now, if $\mathbf{L}$ intersects a 3D line, it can be concluded that this 3D line projects onto the points given by the locus of $\mathbf{q}$, and hence $\mathbf{x}$. Using this principle, we derive the projection matrix.

2.2.2 The Refractive Projection Matrix

Let $\mathbf{A}, \mathbf{B}$ be two homogenous points on some 3D line $\mathbf{L}$. The Plücker coordinates of this line are given as $\mathbf{L}_1 = \begin{pmatrix} A_4\mathbf{B}_{(1,2,3)} - B_4\mathbf{A}_{(1,2,3)} & \mathbf{A}_{(1,2,3)} \times \mathbf{B}_{(1,2,3)} \end{pmatrix}^\top$. The back-projection line $\mathbf{L}$ in Equation 2.7 intersects this line iff these two lines satisfy the Klein quadric constraint [108], defined by the matrix $\mathsf{W} = \begin{pmatrix} 0 & \mathsf{I} \\ \mathsf{I} & 0 \end{pmatrix}$. This can be developed as follows

$$\mathbf{L}_1^\top \mathsf{W} \mathbf{L} = 0 \tag{2.8}$$

$$\begin{pmatrix} \mathbf{L}_{1,(4,5,6)}^\top & \mathbf{L}_{1,(1,2,3)}^\top \end{pmatrix} \mathbf{L} = 0 \tag{2.9}$$

$$\left(\mathbf{L}_{1,(6,1,2)}\right)^\top \begin{pmatrix} -q_3 \\ q_2 t_3 - q_3 t_2 \\ q_3 t_1 - t_3 q_1 \end{pmatrix} \sqrt{(1-\lambda^2)+\lambda^2 q_3^2} = q_3\lambda \left(\mathbf{L}_{1,(4,5,3)}\right)^\top \begin{pmatrix} q_1 \\ q_2 \\ q_2 t_1 - q_1 t_2 \end{pmatrix} \tag{2.10}$$

Squaring both sides removes the square root. Noting that for two vectors $\mathbf{a}$ and $\mathbf{b}$, $(\mathbf{a}\cdot\mathbf{b})^2 = \mathsf{D}\widehat{\mathbf{a}}\cdot\widehat{\mathbf{b}}$ or $\widehat{\mathbf{a}}\cdot\mathsf{D}\widehat{\mathbf{b}}$ where $\mathsf{D} = \mathrm{diag}\begin{pmatrix}1 & 2 & 1 & 2 & 2 & 1\end{pmatrix}$, we get

$$\left(\widehat{\mathbf{S}}(1-\lambda^2+\lambda^2 q_3^2) - \widehat{\mathbf{T}}\lambda^2 q_3^2\right)\mathsf{D}_s\widehat{\mathbf{q}} = 0 \tag{2.11}$$

$$\left(\widehat{\mathbf{S}}(1-\lambda^2) \quad \lambda^2(\widehat{\mathbf{S}}-\widehat{\mathbf{T}})\right)\underbrace{\begin{pmatrix}\mathsf{D}_s & 0\\ 0 & \mathsf{D}_s\end{pmatrix}}_{\bar{\mathsf{D}}_s}\begin{pmatrix}\widehat{\mathbf{q}}\\ q_3^2\widehat{\mathbf{q}}\end{pmatrix} = 0 \tag{2.12}$$

$$\mathbf{S} = \left(\mathbf{L}_{1,(6,1,2)}\right)^\top \underbrace{\begin{pmatrix}0 & 0 & -1\\ 0 & t_3 & -t_2\\ -t_3 & 0 & t_1\end{pmatrix}}_{\mathsf{t}_s}, \mathbf{T} = \left(\mathbf{L}_{1,(4,5,3)}\right)^\top \underbrace{\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ -t_2 & t_1 & 0\end{pmatrix}}_{\mathsf{t}_t} \tag{2.13}$$

$\widehat{\mathbf{S}}$ and $\widehat{\mathbf{T}}$ are defined as functions of t_s and t_t respectively, using Kronecker products and symmetric vectorization as defined in Equation 2.5 (See also [63; 130]).

Observe that equation 2.12. defines a quartic curve in the image coordinates (since the lifted coordinates of $\mathbf{q}$ and $\mathbf{x}$ in Equation 2.5 are related by linear transformations). Note that the camera position $\mathbf{t}_1$, refractive index λ and $\mathbf{R}$ are already absorbed in $\mathbf{q}$. We can thus conclude that a 3D line is imaged to a quartic curve in a perspective camera, if seen through a refractive planar surface. Notice that the coefficents of the quartic curve in $\mathbf{q}$ can be defined as a linear function of $\widehat{\mathbf{L}_1}$. Let

$$\mathbf{c} = \bar{\mathsf{D}}_s^\top\begin{pmatrix}\widehat{\mathbf{S}}^\top(1-\lambda^2)\\ \lambda^2(\widehat{\mathbf{S}}-\widehat{\mathbf{T}})^\top\end{pmatrix} \tag{2.14}$$

$$= \underbrace{\bar{\mathsf{D}}_s^\top\begin{pmatrix}(1-\lambda^2)\mathsf{D}_s{}^{-1}\mathsf{S}_s\mathsf{t}_s\otimes\mathsf{t}_s\mathsf{S}_s{}^\top & 0\\ \lambda^2\mathsf{D}_s{}^{-1}\mathsf{S}_s\mathsf{t}_s\otimes\mathsf{t}_s\mathsf{S}_s{}^\top & -\lambda^2\mathsf{D}_t{}^{-1}\mathsf{S}_t\mathsf{t}_t\otimes\mathsf{t}_t\mathsf{S}_t^\top\end{pmatrix}}_{\mathsf{P}}\begin{pmatrix}\widehat{\mathbf{L}}_{1,(6,1,2)}\\ \widehat{\mathbf{L}}_{1,(4,5,3)}\end{pmatrix} \tag{2.15}$$

Then, equation 2.12 can be written as $\mathbf{c}^\top\begin{pmatrix}\hat{\mathbf{q}}\\ q_3^2\hat{\mathbf{q}}\end{pmatrix} = 0$. $\mathbf{c}$ contains all coefficients of the quartic curve that is the image of $\mathbf{L}_1$. Equation 2.15 shows that this curve can be computed by applying a projection matrix $\mathbf{P}$ to the lifted coordinates of the 3D line.

We have thus derived the projection matrix for a perspective camera viewing a scene separated by

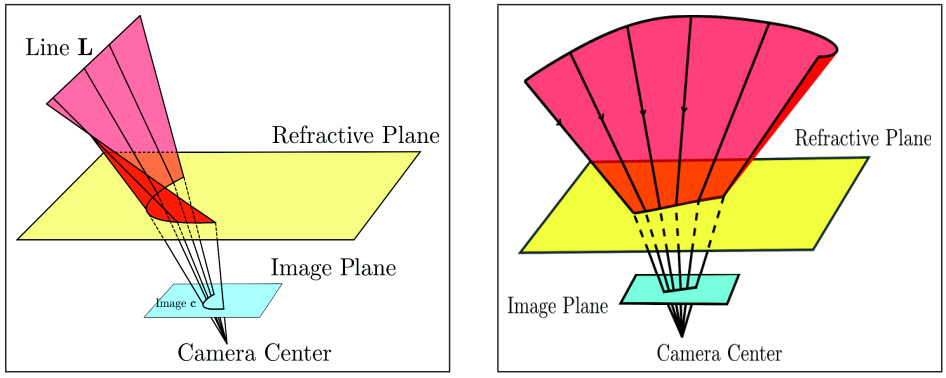


Figure 2.1: (a) shows the projection of a 3D line onto the camera after refraction. (b) shows the back-projection of a line in the image.

a planar refracted surface. It projects the lifted Plücker coordinates of a line, $\mathbf{L}_1$, onto a quartic curve in the image space, whose coefficients are elements of the vector $\mathbf{c}$ (Figure 2.1a).

2.3 The Refractive Fundamental Matrix

In the previous section, we defined the projection matrix for 3D lines. Continuing the same argument, we can see that the epipolar curve in the first image is nothing but the projection of a line. This line, in turn, is the back-projection of an image point from the second camera. Using the back-projection equation 2.7, we get the back-projected ray from the second image as

$$\mathbf{L}_2 = \begin{pmatrix} -q_{2,1}q_{2,3}\lambda \\ -q_{2,2}q_{2,3}\lambda \\ -q_{2,3}\sqrt{1-\lambda^2(1-q_{2,3}^2)} \\ (q_{2,2}t_{2,3} - q_{2,3}t_{2,2})\sqrt{1-\lambda^2(1-q_{2,3}^2)} \\ (q_{2,3}t_{2,1} + t_{2,3}q_{2,1})\sqrt{1-\lambda^2(1-q_{2,3}^2)} \\ (q_{2,1}q_{2,3}t_{2,2} - q_{2,2}q_{2,3}t_{2,1})\lambda \end{pmatrix} \tag{2.16}$$

where $\mathbf{q}_2 = \begin{pmatrix} q_{2,1} & q_{2,2} & q_{2,3} \end{pmatrix}^\top$ and $\mathbf{t}_2 = \begin{pmatrix} t_{2,1} & t_{2,2} & t_{2,3} \end{pmatrix}$ are the image point and the position of the second camera. Substituting $\mathbf{L}_2$ in the place of $\mathbf{L}_1$ in Equation 2.15 will give us the desired result, *i.e.* the image of the back-projection line $\mathbf{L}_2$, which is nothing else than the epipolar curve associated with $\mathbf{q}_2$.

We now express the epipolar constraint, *i.e.* the constraint that a point $\mathbf{q}$ in the first image lies on the epipolar curve of $\mathbf{q}_2$. To do so, we first observe that

$$\widehat{\mathbf{L}_{2,(6,1,2)}} = \mathsf{D}_u^{-1}\mathsf{S}_u\mathsf{t}_u \otimes \mathsf{t}_u\mathsf{S}_u^\top \widehat{\mathbf{q}}_2 q_{2,3}^2\lambda^2, \quad \mathsf{t}_u = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ t_{2,2} & -t_{2,1} & 0 \end{pmatrix} \tag{2.17}$$

$$\widehat{\mathbf{L}_{2,(4,5,3)}} = \mathsf{D}_v^{-1}\mathsf{S}_v\mathsf{t}_v \otimes \mathsf{t}_v\mathsf{S}_v^\top \widehat{\mathbf{q}}_2(1-\lambda^2+q_{2,3}^2\lambda^2), \quad \mathsf{t}_v = \begin{pmatrix} 0 & t_{2,3} & -t_{2,2} \\ t_{2,3} & 0 & t_{2,1} \\ 0 & 0 & -1 \end{pmatrix} \tag{2.18}$$

Now, equation 2.15 can be modified to give the following

$$\begin{pmatrix} \widehat{\mathbf{q}}_2 \\ \widehat{\mathbf{q}}_2 q_{2,3}^2 \end{pmatrix}^\top \begin{pmatrix} 0 & (1-\lambda^2)\Lambda_\mathsf{v} \\ \lambda^2\Lambda_\mathsf{u} & \lambda^2\Lambda_\mathsf{v} \end{pmatrix} \begin{pmatrix} (1-\lambda^2)\Lambda_\mathsf{s} & 0 \\ \lambda^2\Lambda_\mathsf{s} & -\lambda^2\Lambda_\mathsf{t} \end{pmatrix} \bar{\mathsf{D}}_s \begin{pmatrix} \widehat{\mathbf{q}} \\ \widehat{\mathbf{q}} q_3^2 \end{pmatrix} = 0 \tag{2.19}$$

$$\begin{pmatrix} \widehat{\mathbf{q}}_2 \\ \widehat{\mathbf{q}}_2 q_{2,3}^2 \end{pmatrix}^\top \underbrace{\begin{pmatrix} (1-\lambda^2)\lambda^2\Lambda_\mathsf{v}\Lambda_\mathsf{s} & -(1-\lambda^2)\lambda^2\Lambda_\mathsf{v}\Lambda_\mathsf{t} \\ (1-\lambda^2)\lambda^2\Lambda_\mathsf{u}\Lambda_\mathsf{s} + \lambda^4\Lambda_\mathsf{v}\Lambda_\mathsf{s} & -\lambda^4\Lambda_\mathsf{v}\Lambda_\mathsf{t} \end{pmatrix}}_{\mathsf{F}} \bar{\mathsf{D}}_s \begin{pmatrix} \widehat{\mathbf{q}} \\ \widehat{\mathbf{q}} q_3^2 \end{pmatrix} = 0 \tag{2.20}$$

where $\Lambda_i = \mathsf{D}_i^{-1}\mathsf{S}_i\mathsf{t}_i \otimes \mathsf{t}_i\mathsf{S}_i^\top$.

Equation 2.20 defines the Fundamental matrix between two perspective cameras, when looking across a planar refractive surface. F is a matrix of dimensions 12×12 that relates the lifted coordinates in one image to a quartic curve in the other image. In a coordinate system where the refractive plane might be in a general position, this matrix is of dimensions 15×15. It is a function of the relative pose between the cameras, as well as the position of the refractive plane and the refractive index. As expected, the quantity $\mathsf{F}\begin{pmatrix} \widehat{\mathbf{q}} \\ \widehat{\mathbf{q}} q_3^2 \end{pmatrix}$ represents the epipolar quartic curve in the second image.

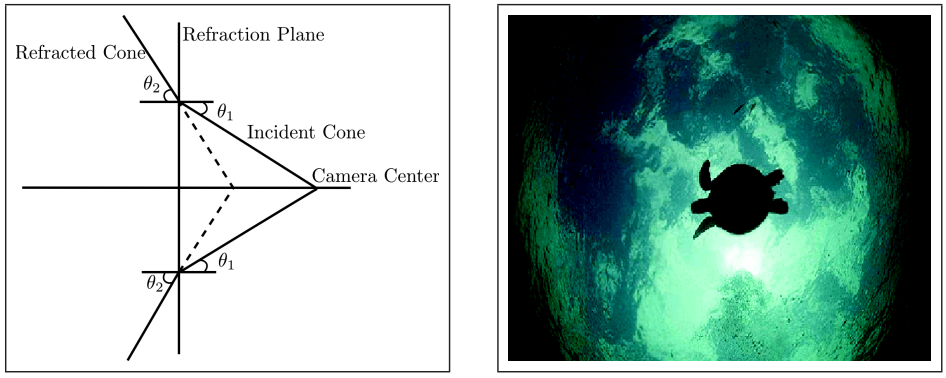


Figure 2.2: (a) shows an illustration of the refraction principle. (b) shows an image of "Snell's Window", a conic that represents the horizon of the outside world. Photo courtesy gerb's photostream, http://www.flickr.com/photos/gerb/196296131/

2.4 The Refractive Homography Matrix

In this section, we derive the relationship between a scene plane in 3D, and its image after refraction. Unlike the traditional case, we show that in the current scenario, this transformation is represented by a family of homographies that map one set of conics to another set.

Consider a cone of rays emerging from the camera, centered at the camera center, with an axis that is aligned with the normal of the refractive plane and an aperture $2\theta_1$. Let us call this the incident cone (Figure 2.2a). Since any ray which is part of this cone surface, makes the same incident angle with the normal of the refractive plane, all refracted rays from this cone make the same angle with the normal of the refractive plane. Thus the refracted rays form a cone centered at the line joining the camera center and the plane, collinear with the normal of the refractive plane, with an aperture of $\theta_2 = 2\sin^{-1}\lambda\sin\theta_1$. Let us call this the refracted cone.

Since the image is formed by cutting the incident cone with the image plane, and the actual scene plane cuts the refracted cone, the 3D homography can be defined as a family of transformations, that transform one conic (formed by the scene plane and refracted cone) to another (formed by the image plane and the incident cone). This family is a function of θ_1.

Given a ray $\mathbf{q}$, the point $\mathbf{Q}$ and the refracted ray direction $\mathbf{R}$ are given by

$$\begin{aligned}
\mathbf{Q} &= (\begin{pmatrix}\mathbf{t}\\1\end{pmatrix}\mathbf{q}^\top - \mathbf{q}\begin{pmatrix}\mathbf{t}^\top & 1\end{pmatrix})\pi && (2.21)\\
&= \begin{pmatrix}(\tilde{\mathbf{q}}\times\mathbf{t})_\times & -\tilde{\mathbf{q}}\\ \tilde{\mathbf{q}}^\top & 0\end{pmatrix}\pi && (2.22)\\
&= \begin{pmatrix}(\tilde{\mathbf{q}}\times\mathbf{t})_\times\mathbf{v} - \tilde{\mathbf{q}}d\\ \cos(\theta_1)\end{pmatrix} = \begin{pmatrix}(\mathbf{v})_\times(\mathbf{t})_\times\tilde{\mathbf{q}} - \tilde{\mathbf{q}}d\\ \cos(\theta_1)\end{pmatrix} && (2.23)\\
\mathbf{R} &= \begin{pmatrix}q_1\lambda & q_2\lambda & \cos(\theta_2) & 0\end{pmatrix}^\top && (2.24)
\end{aligned}$$

where θ_1 is the incident angle and θ_2 is the refracted angle, $\pi = \begin{pmatrix}\mathbf{v}^\top & d\end{pmatrix}^\top$ (general position of the refractive plane) and $(\cdot)_\times$ represents a skew symmetric matrix that corresponds to a cross product with the vector inside parenthesis. The point $\mathbf{S}$ where the above line will strike a plane $\pi_1 = \begin{pmatrix}\mathbf{v}_1^\top & d_1\end{pmatrix}$ is now given by

$$\begin{aligned}
\mathbf{S} &= \left(\mathbf{R}\mathbf{Q}^\top - \mathbf{Q}\mathbf{R}^\top\right)\pi_1 && (2.25)\\
&= \begin{pmatrix}(\mathbf{Q}_{(1,2,3)}\times\mathbf{R}_{(1,2,3)})_\times & \cos(\theta_1)\mathbf{R}_{(1,2,3)}\\ -\cos(\theta_1)\mathbf{R}_{(1,2,3)} & 0\end{pmatrix}\pi_1 && (2.26)\\
&= \begin{pmatrix}-(\mathbf{v}_1)_\times(\mathbf{Q}_{(1,2,3)}\times\mathbf{R}_{(1,2,3)}) + \cos(\theta_1)\mathbf{R}_{(1,2,3)}d_1\\ -\cos(\theta_1)\mathbf{R}_{(1,2,3)}\cdot\mathbf{v}_1\end{pmatrix} && (2.27)\\
&= \begin{pmatrix}-(\mathbf{v}_1)_\times(\mathbf{v})_\times(\mathbf{t})_\times\tilde{\mathbf{q}}\times\mathbf{R}_{(1,2,3)} + (\mathbf{v}_1)_\times(\tilde{\mathbf{q}}\times\mathbf{R}_{(1,2,3)})d + \cos(\theta_1)\mathbf{R}_{(1,2,3)}d_1\\ -\cos(\theta_1)\mathbf{R}_{(1,2,3)}\cdot\mathbf{v}_1\end{pmatrix} && (2.28)\\
&= \begin{pmatrix}-(\mathbf{v}_1)_\times(\mathbf{v})_\times(\mathbf{t})_\times + d\begin{pmatrix}\mathbf{v}_1\end{pmatrix}_\times & d_1\mathrm{I}_{3\times3}\\ \begin{pmatrix}0 & 0 & 0\end{pmatrix} & -\mathbf{v}_1^\top\end{pmatrix}\begin{pmatrix}\tilde{\mathbf{q}}\times\mathbf{R}_{(1,2,3)}\\ \cos(\theta_1)\mathbf{R}_{(1,2,3)}\end{pmatrix} && (2.29)
\end{aligned}$$

The vector $\tilde{\mathbf{q}}\times\mathbf{R}_{(1,2,3)}$ expands to give $\begin{pmatrix}q_2(\cos(\theta_2)-\cos(\theta_1)) & q_1(\cos(\theta_1)-\cos(\theta_2)) & 0\end{pmatrix}^\top$. We can

isolate the cosines into a separate matrix to get

$$\mathbf{S} = \begin{pmatrix} -(\mathbf{v}_1)_\times(\mathbf{v})_\times(\mathbf{t})_\times + d\left(\mathbf{v}_1\right)_\times & d_1\mathsf{I}_{3\times3} \\ \begin{pmatrix}0 & 0 & 0\end{pmatrix} & -\mathbf{v}_1^\top \end{pmatrix} \begin{pmatrix} q_2(\cos(\theta_2)-\cos(\theta_1)) \\ q_1(\cos(\theta_1)-\cos(\theta_2)) \\ 0 \\ \cos(\theta_1)q_1\lambda \\ \cos(\theta_1)q_2\lambda \\ \cos(\theta_1)\cos(\theta_2) \end{pmatrix} \tag{2.30}$$

$$= \begin{pmatrix} -(\mathbf{v}_1)_\times(\mathbf{v})_\times(\mathbf{t})_\times + d\left(\mathbf{v}_1\right)_\times & d_1\mathsf{I}_{3\times3} \\ \begin{pmatrix}0 & 0 & 0\end{pmatrix} & -\mathbf{v}_1^\top \end{pmatrix} \begin{pmatrix} 0 & a & 0 \\ -a & 0 & 0 \\ 0 & 0 & 0 \\ \lambda\cos(\theta_1) & 0 & 0 \\ 0 & \lambda\cos(\theta_1) & 0 \\ 0 & 0 & \cos(\theta_2) \end{pmatrix} \tilde{\mathbf{q}} \tag{2.31}$$

$$= \begin{pmatrix} -(\mathbf{v}_1)_\times(\mathbf{v})_\times(\mathbf{t})_\times\left(\mathbf{a}\right)_\times + d\left(\mathbf{v}_1\right)_\times\left(\mathbf{a}\right)_\times + d_1\begin{pmatrix}\lambda\cos(\theta_1)\mathsf{I}_{2\times2} & \mathbf{0} \\ \mathbf{0}^\top & \cos(\theta_2)\end{pmatrix} \\ -\mathbf{v}_1^\top\begin{pmatrix}\lambda\cos(\theta_1)\mathsf{I}_{2\times2} & \mathbf{0} \\ \mathbf{0}^\top & \cos(\theta_2)\end{pmatrix} \end{pmatrix} \tilde{\mathbf{q}} \tag{2.32}$$

$$\tag{2.33}$$

$$\sim \mathbf{H}_{\theta_1}\mathbf{x} \tag{2.34}$$

with

$$\mathbf{H}_{\theta_1} = \begin{pmatrix} -(\mathbf{v}_1)_\times(\mathbf{v})_\times(\mathbf{t})_\times\left(\mathbf{a}\right)_\times + d\left(\mathbf{v}_1\right)_\times\left(\mathbf{a}\right)_\times + d_1\begin{pmatrix}\lambda\cos(\theta_1)\mathsf{I}_{2\times2} & \mathbf{0} \\ \mathbf{0}^\top & \cos(\theta_2)\end{pmatrix} \\ -\mathbf{v}_1^\top\begin{pmatrix}\lambda\cos(\theta_1)\mathsf{I}_{2\times2} & \mathbf{0} \\ \mathbf{0}^\top & \cos(\theta_2)\end{pmatrix} \end{pmatrix} \mathsf{R}^\top \tag{2.35}$$

$$a = \cos(\theta_2)-\cos(\theta_1) \tag{2.36}$$

$$\mathbf{a} = \begin{pmatrix}0 & 0 & a\end{pmatrix}^\top \tag{2.37}$$

Equation 2.35 shows how the homography matrix can be expressed as a function of the incident and refracted angles. From the above expression, it is clear that there exists a *family* of homographies that are linear in the cosines of these angles, which fully represent the projection of a plane onto the image. Although the above expression can be further developed to obtain an expression in lifted coordinates that is independent of these angles, i.e., a homography that applies to all points of the image and scene plane, we omit the derivation here because of space constraints.

2.5 Exploiting Snell's Window

Till now, we have discussed the general case of planar refraction, which is applicable to any scene irrespective of the relative refractive index. Of particular importance, is a special case, when the camera is in a denser medium and looking outward into a lighter medium. This amounts to a relative refractive index that is > 1 (as per our current convention). In this case, it is possible to image the *peripheral rays*: rays that make an angle of $\pi/2$ with the refractive surface normal after refraction, and hence are parallel to the refracting surface. The importance of these rays is that any ray with an incident angle greater than the incident angle of these rays is *reflected* back into the denser medium. This set of rays thus captures the periphery of the world on the other side of the surface, which in the case of underwater imagery is the horizon (Figure 2.2b).

2.5.1 The Image of Snell's Window

Consider a point $\mathbf{x}$ that is on the image of this periphery. Since after refraction the ray from this point is parallel to the refractive surface, the refracted angle is $\frac{\pi}{2}$, which means

$$1 - \lambda^2 + \lambda^2(\pi^\top \mathbf{q})^2 = 0 \tag{2.38}$$

where $\mathbf{q} = \begin{pmatrix} (\mathsf{R}^\top \tilde{\mathbf{x}})^\top & 0 \end{pmatrix}^\top$, with $\tilde{\mathbf{x}}$ being the unit vector of image point $\mathbf{x}$ in the camera coordinate system. Let us develop these equations further.

$$\begin{aligned}
\mathbf{q}^\top \pi \pi^\top \mathbf{q} &= 1 - \frac{1}{\lambda^2} && (2.39)\\
\tilde{\mathbf{x}}^\top \mathsf{R}^\top \mathbf{v}\mathbf{v}^\top \mathsf{R} \tilde{\mathbf{x}} &= 1 - \frac{1}{\lambda^2} && (2.40)\\
\frac{1}{\|\mathbf{x}^2\|} \mathbf{x}^\top \mathsf{R}^\top \mathbf{v}\mathbf{v}^\top \mathsf{R} \mathbf{x} &= 1 - \frac{1}{\lambda^2} && (2.41)\\
\mathbf{x}^\top \mathsf{R}^\top \mathbf{v}\mathbf{v}^\top \mathsf{R} \mathbf{x} &= \mathbf{x}^\top \mathsf{I}_{3\times 3} \mathbf{x} (1 - \frac{1}{\lambda^2}) && (2.42)\\
\mathbf{x}^\top \left(\mathsf{R}^\top \mathbf{v}\mathbf{v}^\top \mathsf{R} - (1 - \tfrac{1}{\lambda^2}) \mathsf{I}_{3\times 3} \right) \mathbf{x} &= 0 && (2.43)
\end{aligned}$$

which shows that the periphery is a conic in image coordinates. The term $\mathsf{R}^\top \mathsf{v}$ represents the refractive plane normal in the camera coordinate system. One of the main advantages of Equation 2.43 is that one of the conic's eigenvectors is the normal of the refractive plane.

$$\left(\mathsf{R}^\top \mathbf{v}\mathbf{v}^\top \mathsf{R} - (1 - \tfrac{1}{\lambda^2}) \mathsf{I}_{3\times 3} \right) \left(\mathsf{R}^\top \mathbf{v} \right) = \frac{1}{\lambda^2} \mathsf{R}^\top \mathsf{v} \qquad (2.44)$$

since $\mathbf{v}^\top \mathbf{v} = 1$.

In fact, it is easy to show that the eigenvalues of the above matrix are $\frac{1}{\lambda^2}, -1 + \frac{1}{\lambda^2}, -1 + \frac{1}{\lambda^2}$. Thus the important property of the eigenvalues and eigenvectors of the above matrix is that *the only positive eigenvalue is the square inverse of the relative refractive index, while the corresponding eigenvector is the normal of the refractive plane in the camera coordinate system*. Thus we have shown that the image of the horizon after refraction from a planar scene is a conic, and that using a simple SVD based algorithm it is possible to recover both the plane normal and the relative refractive index from a *single* image.

2.5.2 Relative Pose Computation

Given that we can estimate the plane normal (in the camera coordinate system), and the refractive index, we now show how to compute the relative pose of a second camera with respect to the first. Let us first note that since both the cameras see the same refractive surface, estimating the normal of

the plane in the camera coordinate systems of both cameras can help estimating the rotation between them upto a rotation *about* the normal itself. Additionally, we can set the perpendicular distance from the camera center of one of the cameras to the refractive plane to 1, so as to set the scale of the scene. Since we have chosen a world coordinate system where the plane normal is aligned with the Z-axis, there are 4 unknowns that remain to be computed, namely the rotation about the Z-axis and the translation parameters.

In order to solve for these parameters, let us start by observing the fundamental matrix equation (2.20). Observe that the image points (rays) are a function of the unknown rotation matrix of the second camera, which is of the form $\mathsf{R} = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) & 0 \\ -\sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Thus, the vector $\begin{pmatrix} \widehat{\mathbf{q}}_2 & \widehat{\mathbf{q}}_2 q_{2,3}^2 \end{pmatrix}^\top$ is quadratic in the elements of the rotation matrix R. Equation (2.20) thus gives constraints on α and $\mathbf{t}_2$. Thus, 4 point-correspondences should be sufficient for a solution, although we have not been able to obtain a solution using either minimal solution packages or empirically.

2.6 Discussion and Conclusion

In this chapter, we have defined the epipolar geometry for scenes where a single plane separates two media with different refractive indices. The camera(s) are in one media, while the object(s) being observed are on the other side. Such scenarios occur often in underwater vision. We have shown that for the case when the camera is in a denser medium, it is even possible to estimate the refractive index and the refractive surface geometry from a *single* image, by observing the so-called "Snell's Window". Further, we have shown a method to compute the relative pose between two cameras in such a scenario.

The contributions of this work are theoretical, and of conceptual value. Our main motivation for this work has been to explicitly model the geometry behind imaging in such scenarios. Apart from the theoretical value of the results we have shown thus far, what is also important is that using the results of this chapter, it is actually possible now to *measure* the deviation of the standard perspective model in areas like underwater imagery. This is of importance while modeling errors in observation using a perspective model.

3

Photogeometric Reconstruction of Transparent Objects

3.1 Introduction

In the previous chapter we looked at the multiple view geometry associated with a plane. While we presented approaches that recover the geometry of the scene and camera using a purely geometric fashion, a practical reconstruction algorithm has to contend with scene illumination effects also in dealing with transparent refractive surfaces. Also while we derived multiview constraints relating to a plane modeling transparent surfaces in a piecewise planar approach using the constraints we derived is not easily achieved. To this extent, in this chapter we try to derive an image based approach to reconstruct general reflective and transparent surfaces that combines both geometric and photometric

information.

Contents

Image-based reconstruction of transparent objects has gathered interest in the last few years [25; 64; 73; 88; 89; 91; 93]. Several cues and approaches have been proposed for this task. They could be broadly classified as approaches that rely on physical (material) properties of transparent objects, and approaches that try to extend traditional shape acquisition approaches to the case of transparent objects. Among the approaches relying on material properties, geometric and photometric cues are the most prominent inputs to reconstruction algorithms.

Transparent objects referred to in the recent computer vision literature could be categorized into two kinds: specular and transluscent. Specular transparent objects are those whose surfaces exhibit specular reflection and refraction: for every ray of light incident on the surface, one refracted ray and one reflected ray is emitted. Transluscent surfaces generally do not observe this property, they emit multiple reflected/refracted rays for every incident ray. In this chapter, we focus on specular

transparent objects. For such objects, geometric cues are an important source of information. A popular type of approach is based on the so-called light-path triangulation principle [73]. Roughly speaking, one acquires images of a reference object with known shape, seen through the transparent object or reflected by it. Using camera calibration information, known relative position(s) of reference object and camera, and matches between the reference object and the camera images, one may recover the "light paths" associated with such matches by techniques akin to structure-from-motion and multi-view geometry, provided sufficiently many images in different positions are acquired. A light path refers here to a sequence of straight line segments that connect an object point and its image, where successive segments are the result of reflections/refractions in the transparent object's surface(s). Let us consider an object point emitting light in "all" directions. Underlying the light path triangulation approach is the assumption that only light emitted in a single direction, eventually hits the image plane of the camera, after undergoing reflections/refractions in the transparent object. This in turn relies on assuming an infinitesimal camera aperture. Also, in some cases there may be multiple light paths associated with a single object point, but it is assumed that a single one is observable by the camera.

Similar works use specular highlights produced by transparent/specular objects and the geometric cues these provide on the object surface. While specular highlights are sometimes referred to as photometric cues, we rather consider them as geometric ones – photometric analysis is usually restricted to identifying highlights, after which only geometric information is used. This is similar to light path triangulation and similar methods, where photometric information is used to determine matches but is then discarded from further processing.

In our work, we fully combine photometric and geometric information for the reconstruction of transparent objects. By photometric information, we mean the intensity values of image pixels and the irradiance values of points of the reference object whose reflections/refractions are imaged by the camera (in practice, we use screens as reference objects). Let us come back to light paths, as defined just above. In addition to modeling the *geometric* path light follows from an object point until its image point, we also model the *photometric* changes occurring along this path. In particular, real objects are neither perfect mirrors nor perfectly transparent: hence, each reflection/refraction causes a loss of irradiance. More precisely, light hitting a refractice surface gets partly reflected, partly refracted. For specific surface materials, theoretic models exist for how much light gets reflected and how much refracted. This depends on the refractive indices of the media in which the light

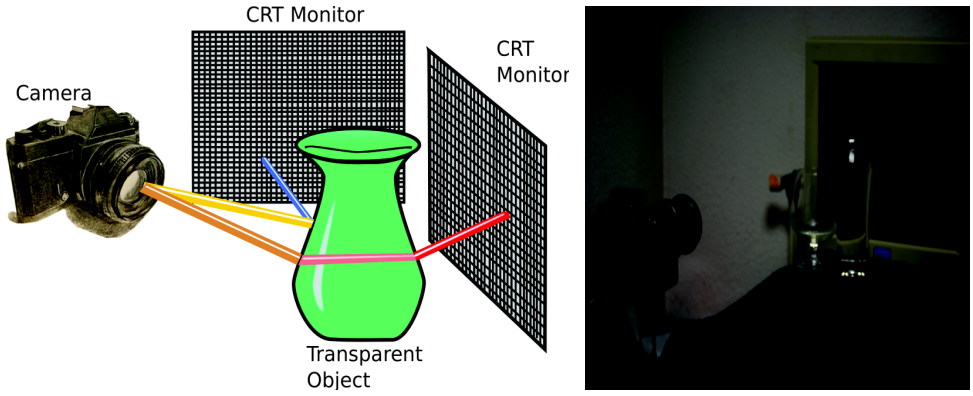


Figure 3.1: (a) Description of the general theory behind our approach. While the acquisition is similar to that of Kutulakos *et al* [73], we also include radiance in our measurements (depicted by the changing color of rays while they travel from the illuminant to the camera pixel). (b) Our setup to "acquire" the shape of transparent objects consists simply of a CRT monitor as light source and a camera looking at light reflected / refracted off the object.

travels as well as on the incidence angle between light path and surface normal. Hence, modeling and measuring this phenomenon brings about information on the shape of the transparent object; importantly, this information is complementary to the information given by purely geometric input to light path triangulation. In essence: consider a scenario where the available information (number of images) is insufficient to provide a unique solution for the surface shape using light path triangulation. In other words, there are multiple surface shapes that reflect/refract light rays emitted from object points in a way that they hit the matching image points. However, when tracing light paths and in addition, tracing the loss of irradiance along the paths due to reflection/refraction and comparing this against measured image intensities, then one gets one additional constraint per match that may allow to uncover a unique solution for our reconstruction problem.

We present an approach that combines the merits of utilizing both geometric and photometric cues. Our approach is along the same lines as [73], with one important difference. For every "light-path" that we capture, we record *both* geometric information (position and direction of light rays captured by and originating from light source, depending on requirement) *and* radiometric information (radiance of light at the beginning and end of each light-path). We show that our approach benefits from the following advantages because it uses a combination of cues:

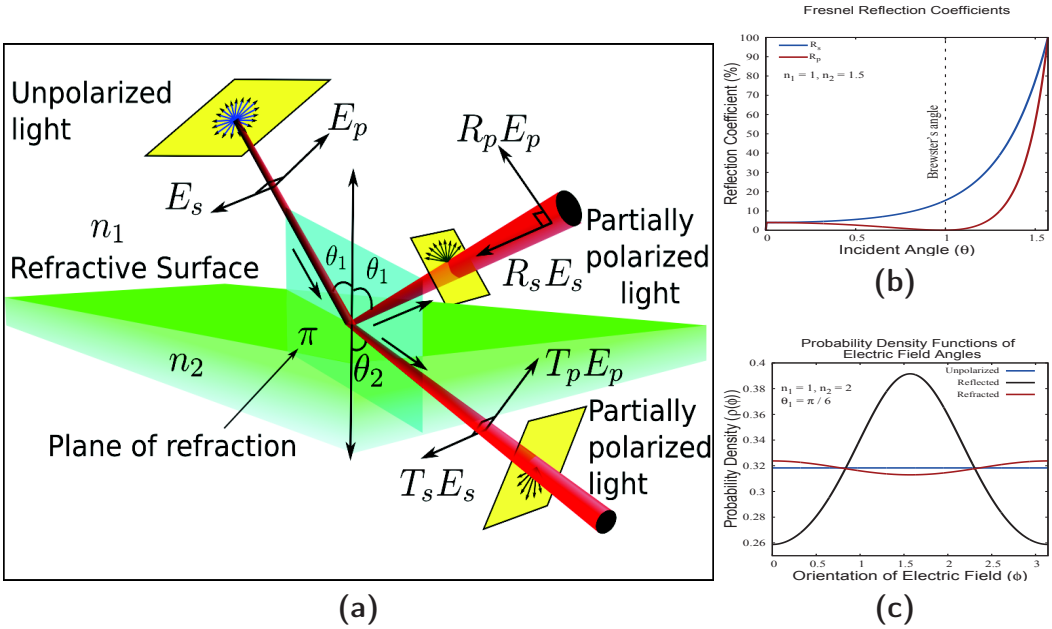


Figure 3.2: (a) Depiction of a the phenomenon of specular reflection and refraction. A single ray of light incident on the surface of a transparent object is partly reflected and partly refracted. Both rays contain a fraction of energy of the incident ray, but different polarizations. (b) Reflection coefficients computed using Fresnel equations. Courtesy [36] (c) Probabilities of orientation of electric fields on incidence (blue, unpolarized), after reflection (black) and refraction (red) off a surface. Unpolarized light becomes partially polarized after a single bounce. Angles are computed in a plane perpendicular to direction of ray propagation.

1. It is able to extend the minimal case scenarios of reconstructions beyond those proposed in Kutulakos *et al.* [73]. Thus we reduce the number of views required for reconstruction.

2. We show that it is possible to reconstruct objects in the presence of significant amounts of inter-reflections, caustics and other phenomena that would normally prove a hindrance to geometry based approaches.

The rest of the chapter is structured as follows. Section 3.2 places previous works in perspective. Section 3.3 outlines the underlying theory and minimal solutions for our approach, and discuss some practical problems to be countered. Section 3.6 shows results for various cases. Finally, Section 3.7 presents an overall summary of our approach, points to future directions of research and presents potential applications.

Notation: In the rest of the chapter, we use the following convention. Bold symbols like $\mathbf{E}$ denote vectors with direction and magnitude, while $\hat{\mathbf{E}}$ denotes its corresponding (unit) direction and E, its magnitude. Subscripts $\mathbf{E}_s$ are either used to denote the components of a vector in a corresponding direction (in which case E_s is the magnitude of the component), or in case of scalars, used to differentiate variables referring to the same physical phenomenon or variable.

3.2 Related Work

In the past, several approaches have used either geometric or photometric cues to reconstruct transparent surfaces. Geometric approaches typically measure the deviation from perspective imaging produced by a refractive transparent object, and recover the shape as a solution that explains the observation. In [73], the authors present a minimal case study of the conditions in which refractive surfaces can be reconstructed. They re-cast transparent object reconstruction as the task of reconstructing the path of each individual ray of light that is captured by a camera after refraction through a transparent surface. They show that individually, it is impossible to reconstruct the entire trajectory of a ray of light after it has undergone more than 2 refractions. Earlier approaches have focussed on parameterizing the object to be reconstructed and then obtaining its parameters by explaining the distortion it produces in several cameras [9]. Other examples of shape recovery from distortion analysis include the more recent work by [91], which analyzes the specific case of a single dynamic transparent surface that distorts a known background and is observed by multiple cameras. The object surface is reconstructed using a modified stereo reconstruction approach that uses a new distance function. Finally, some recent approaches have also looked at learning based solutions that minimize a certain criterion of optimality [32; 140].

Apart from geometry, photometric information also turns out to be very important in the case of transparent objects since they simultaneously reflect and refract light. In fact, since they reflect light like a specular surface, in principle, any method for specular object reconstruction (like traditional photometric stereo) may be applied. Many recent photometric approaches have tried to reconstruct transparent surfaces by studying their specularities. While [148] provides a low cost approach to reconstruction by studying specular highlights, [91] shows how to reconstruct transparent surfaces with inhomogeneous refractive interiors, by measuring highlights multiple times to remove extraneous effects like scattering, interreflections etc. One important aspect of photometry is polarization.

When unpolarized light is reflected or transmitted across a refractive surface, it gets partially polarized. This degree of polarization is connected to the shape of the object, and several authors have attempted to explore this connection. In [65] the authors measure the polarization angle using multiple images from a single view taken with different orientations of a polarizer. They then use Fresnel theory to derive a relationship between measured polarization state and the angle of reflection. Another approach involving polarization is presented in [88]. Here, the authors start from an initial guess for the shape of the object involved, and using a technique called polarization ray tracing, they predict the polarization information for that shape, and arrive at the correct shape by minimizing the difference between observed and predicted states.

Since refractive objects present a challenging reconstruction problem, many authors have resorted to using active approaches for reconstructions. Methods like fluorescent immersion range scanning [64] and tomographic reconstructions present alternate approaches that are of practical value when objects are of manageable size.

3.3 Physical Modeling

Consider a ray of light traveling in space in a direction represented by the vector $\hat{\mathbf{i}}$. Suppose this ray strikes a refractive medium as shown in Figure 3.2a (inset). This ray is accompanied by an electric field, whose oscillation direction we denote by the vector $\hat{\mathbf{E}}$. The radiant energy of this ray is related to the magnitude of its Poynting vector [60], which is also proportional to the square of the maximum amplitude of its electric field

$$I_{in} \propto E^2 \tag{3.1}$$

where E denotes the maximum amplitude of the electric field [60]. When such a ray falls on the boundary of a refractive medium, part of this energy is transmitted, and the rest is reflected (assuming no absorption by the media). Our contention is that the ratio of irradiance of the transmitted or reflected rays with respect to the incident ray contains information about the 3D structure of the surface. The ratio of these energies can be computed using Fresnel equations relative to the *plane of refraction*, the plane that contains the incident ray, the normal and the refracted ray.

Let us denote the normal of this plane as $\hat{\pi}$, and subsequently divide the electric field into two components (directions): one that is perpendicular to this plane ($\hat{\mathbf{E}}_\mathbf{s}$) and the other that is parallel to

this plane ($\hat{\mathbf{E}}_\mathbf{p}$). ($\therefore E^2 = E_s^2 + E_p^2$). Thus given incident angle θ_1 and refracted angle θ_2 (Figure 3.2a (inset)), and refractive indices (n_1, n_2), the following equations describe the phenomenon

$$
\begin{aligned}
n_1 \sin(\theta_1) &= n_2 \sin(\theta_2) && (3.2)\\
R_s &= \frac{\sin^2(\theta_1 - \theta_2)}{\sin^2(\theta_1 + \theta_2)} && (3.3)\\
R_p &= \frac{\tan^2(\theta_1 - \theta_2)}{\tan^2(\theta_1 + \theta_2)} && (3.4)\\
T_s &= \frac{4\sin^2(\theta_2)\cos^2(\theta_1)}{\sin^2(\theta_1 + \theta_2)} && (3.5)\\
T_p &= \frac{4\sin^2(\theta_2)\cos^2(\theta_1)}{\sin^2(\theta_1 + \theta_2)\cos^2(\theta_1 - \theta_2)} && (3.6)\\
I_{rl} &\propto E_s^2 R_s + E_p^2 R_p && (3.7)\\
I_{rr} &\propto E_s^2 T_s + E_p^2 T_p && (3.8)
\end{aligned}
$$

Equations (3.3)-(3.4) are called Fresnel equations [36; 60].

Polarization A single light ray's electric field could be oriented in any direction that is perpendicular to the direction of ray propagation. The above equations show that transmitted and reflected energies depend on this direction. However, since a camera is a light integration device, it records information from several such rays in a range of directions at each pixel. This helps us *integrate out* the bias of a single ray towards a particular direction. However, we assume that the illuminant that we use generates light rays that are not biased towards any electric field orientation (*unpolarized* light).

For unpolarized light, the values of E_s and E_p can be written as $E_s = E\cos(\phi)$ and $E_p = E\sin(\phi)$, where ϕ is the angle made by the $\mathbf{E}$ vector with $\hat{\pi}$. We use the term *polarization angle* to refer to ϕ in this chapter. In the next section, we analyze the change in irradiance captured by a camera when light bounces (reflects or refracts) off a surface *once* and *twice*, since it is known that geometrically it is impossible to reconstruct more than two bounces, when reconstructing each light-path individually [73]. With photometry too, reconstruction of light-paths with two bounces is not possible without using Fresnel equations.

I	Image Intensity.
I_l	Intensity after reflection.
I_r	Intensity after refraction.
${}^3I_{lrl}$	Image intensity after 3 bounces. (reflection, refractin, reflection).
${}^kI_{llr\ldots r}$	Intensity after k bounces of reflections and/or refractions.
$\rho(\phi)$	Polarization angle probability.
	Single Bounce Notations
π	Plane of refraction.
θ_1	Angle of incidence.
θ_2	Angle of reflection/refraction, depending on case.
$\mathbf{E}$	Electric field vector accompanying light ray.
E	Electric field magnitude accompanying light ray. $\|\|\mathbf{E}\|\|$
E_s, E_p	Magnitude of components of $\mathbf{E}$ perpendicular and parallel to plane.
ϕ	Angle made by electric field with π.
R_s, T_s	Reflection/Transmission coefficient for $\mathbf{E}$ along the vector π.
R_p, T_p	Reflection/Transmission coefficient for $\mathbf{E}$ perpendicular to π.
	Note that $E_s = E\cos(\phi), E_p = E\sin(\phi)$

Table 3.1: Single Bounce Notations

Double Bounce Notations (First Bounce)	
${}^1\pi$	Plane of refraction.
${}^1\theta_1$	Angle of incidence.
${}^1\theta_2$	Angle of reflection/refraction, depending on case.
${}^1\mathbf{E}$	Electric field vector accompanying light ray.
${}^1E$	Electric field magnitude accompanying light ray.
${}^1\phi$	Angle made by electric field with π^1.
${}^1R_s, {}^1T_s$	Reflection/Transmission coefficient for $\mathbf{E}^1$ along vector π^1
${}^1R_p, {}^1T_p$	Reflection/Transmission coefficient for $\mathbf{E}^1$ perpendicular to π^1
Double Bounce Notations (Second Bounce)	
${}^2\pi$	Plane of refraction.
${}^2\theta_1$	Angle of incidence.
${}^2\theta_2$	Angle of reflection/refraction, depending on case.
${}^2\mathbf{E}$	Electric field vector accompanying light ray.
${}^2E$	Electric field magnitude accompanying light ray.
${}^2\phi$	Angle made by electric field with π^2.
${}^2R_s, {}^2T_s$	Reflection/Transmission coefficient for $\mathbf{E}^2$ along vector π^2
${}^2R_p, {}^2T_p$	Reflection/Transmission coefficient for $\mathbf{E}^2$ perpendicular to π^2

Table 3.2: Double Bounce Notations

3.4 Theory of Bounces

In this section, we develop relations between reflected or transmitted light energy and its relation to the 3D structure of the transparent object. For convenience, we hereafter denote an ensemble of rays traveling in the same direction in the same position in space but with different polarization angles, as a *single ray*. We further use the symbol I to represent the irradiance of such an ensemble as measured by a camera. All notations used in this chapter, are summarized in tables 3.1 and 3.2.

3.4.1 Case 1: Single Bounce

Consider a ray $\hat{\mathbf{i}}$ back-projected from a CCD pixel of a camera. Assume that this back-projected ray reflects off a transparent surface *once*, before hitting any number of illuminant points $\mathtt{X}_m$ (one illuminant per image) in space. Figure (reference) depicts this scenario. In such a case, the *average* irradiance of reflected components can be expressed as

$$
\begin{aligned}
I_{rl} &= R_s E^2 \frac{1}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^2(\phi) d\phi + R_p E^2 \frac{1}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^2(\phi) && (3.9)\\
&= \frac{1}{2}(R_s E^2 + R_p E^2) && (3.10)\\
&= \frac{R_s + R_p}{2}(E^2) && (3.11)\\
&= \frac{R_s + R_p}{2} I_{in} && (3.12)
\end{aligned}
$$

where I_{in} represents the radiance of the ray emitted from one point $\mathtt{X}_m$. R_s and R_p are defined in equations 3.3 and 3.4. If the light is partially polarized, then ϕ varies within a range (l, u) depending on the degree of polarization. Also, since all directions are not present in equal probability, the irradiance equation has to be modified to

$$
I_{rl} = R_s E^2 \int_l^u \cos^2(\phi)\rho(\phi) d\phi + R_p E^2 \int_l^u \sin^2(\phi)\rho(\phi) \qquad (3.13)
$$

where ϕ is the probability distribution function with $\int_l^u \rho(\phi) = 1$. The transmitted energy I_{rr} also has a similar formulation, following from equation (3.8). Notice that equations (3.3)-(3.4) can be used to show that light rays after reflection / refraction are partially polarized [60]. We now list the two main triangulation results of the single bounce case. The problem of triangulation is to estimate the point(s) at which every light-path intersects the transparent surface along with the normal at that point, given 2D-3D correspondences. We use the notation of [73] to represent the triangulation problems ($< 1, 2, 3 >$ means 1 view, 2 bounces and 3 3D correspondences per light-path). Both the problems mentioned below have been earlier shown to be tractable with lesser information [73].

Result 1: $< 1, 1, 1 >$ **Triangulation** Consider any point at depth d_k from the camera center along $\hat{\mathbf{i}}$. This point is associated with a unique normal $\hat{\mathbf{n}}_k$ such that the reflected ray passes through the point $\mathtt{X}_m$. In the case of refraction, this point is associated with a set of normal-refractive index pairs that allow the refracted ray to pass through $\mathtt{X}_m$. However, it is *also* associated with a radiance ratio $I_\alpha = I_{rl}/I_{in} = \frac{2}{R_s+R_p}$ (or $I_\alpha = I_{rr}/I_{in}$) which is a function of the reflection (refraction) angle and the relative refractive index $\frac{n_1}{n_2}$. Thus when the refractive index is known we have one variable d_k and one constraint I_α. Thus, $< 1, 1, 1 >$ *Triangulation is tractable.*

Result 2: $< 1, 1, 2 >$ **Triangulation** In this case, we need two observations $(\mathtt{X}_1, \mathtt{X}_2)$ to compute the value of the incident ray direction, similar to [73]. Unlike them however, even when light is only *reflected* off a transparent object surface, equation (3.12) can be used to solve for relative refractive index $\frac{n_1}{n_2}$.

3.4.2 Case 2: Double Bounce

When light bounces off a transparent dielectric surface, it is partially polarized. If the outgoing light's electric field is parametrized by the angle ψ then we have

$$\rho(\psi)_l = \frac{\sqrt{R_s R_p}}{|(R_p - R_s)\cos^2(\psi) + R_s|}\frac{1}{\pi} \tag{3.14}$$

$$\rho(\psi)_r = \frac{\sqrt{T_s T_p}}{|(T_p - T_s)\cos^2(\psi) + T_s|}\frac{1}{\pi} \tag{3.15}$$

where the limits are ($l = -\frac{\pi}{2}, u = \frac{\pi}{2}$) (See Appendix (Section .2). Thus, in the case of a second bounce, the incident light is already partially polarized. In this case, we use equation 3.13 to derive the irradiance measured as a function of the radiance emitted by the illuminant. This irradiance is given as

$$ {}^2I_{ll} = E^2 \frac{1}{2} \left(({}^2R_s \ {}^1R_p + {}^1R_s \ {}^2R_p) \sin^2(\alpha) + ({}^1R_s \ {}^2R_s + {}^1R_p \ {}^2R_p) \cos^2(\alpha) \right) \tag{3.16} $$

where α is the angle between the normals to the two planes of refraction/reflection. (Please refer to the Appendix (Section .2) for a proof).

Result 3: $< 2, 2, 2 >$ **Triangulation** Consider two cameras looking at a transparent object, which refracts light from a known illuminant twice. Further assume that for each camera, two measurements were made per light ray. This corresponds to the case $< 2, 2, 2 >$ as per the convention of [73]. Since equations (3.14, 3.15) give us one extra constraint per light ray, it is now possible to solve for the 3D structure of the transparent object given known refractive index.

Result 4: $< 3, 2, 2 >$ **Triangulation** It is known that this problem is tractable in the case of known refractive index. However, it is now possible to solve for this case even with unknown refractive index, since all the radiance ratios obtained have to be consistent with both the light-path geometry and photometric observations.

A table containing the minimal scenarios for transparent objects that can now be solved with the help of irradiance measurements is shown in Table 3.3. It is interesting to note that transparent objects have *lesser* minimal requirements for reconstruction than mirror like objects. While in this section we have presented a counting argument to solving single and double bounce cases, we will detail the characteristics of the solutions that can be obtained in the next section. Specifically, we analyze the solution space for both cases and present some practical aspects of data acquisition.

3.5 Methods of Solutions

Understanding the physics behind the radiance of a single ray of light captured by a camera gives us the radiometric tools needed to reconstruct a light-path. In this section, we use this understanding

One reference point. ($M = 1$)

	$K = 1$	$K = 2$	$K \geq 3$
$N = 1$	$\surd$		
$N = 2$	$\surd$ $\times^*$		

Two or more reference points. ($M \geq 1$)

	$K = 1$	$K = 2$	$K \geq 3$
$N = 1$	$\surd$ $\times^*$		
$N = 2$	$\surd$ $\times^*$	$\surd$	
$N = 3$	$\surd$ $\times^*$	$\surd$ $\times$	
$N \geq 4$	$\surd$ $\times^*$	$\surd$ $\times^*$	

Table 3.3: Tractable triangulation problems. Updated from [73]. The asterik symbol represents the fact that even in the case of only reflection, the *relative refracitve index* can be computed.

to derive the solution space of depths and normals that could explain the geometric and photometric characteristics of a light-ray observed by a camera.

3.5.1 Single Reflection or Refraction

The category of single bounce cases pertains to the scenario where a light-path is composed of two line segments. This could mean that the light ray associated with the light-path underwent refraction or reflection. In this section, we deal with the case of reflections, i.e. the camera acquires images of the monitor, reflected in the object to be reconstructed. The case of refraction is analogous.

Let us now consider a single pixel and the acquired intensity I_l. The pixel's line of sight is known by calibration. Let d be the depth of the object along this line of sight, P be the intersection point of the object surface and the line of sight, and n the surface normal at that point. Further, given the matching (cf. above), we know the point X on the monitor whose reflection is seen in the pixel.

Our goal is to compute the depth d. We do so by first computing the incident angle θ_1 between the surface normal and the incident light ray, from which it is trivial to compute d.

Since our setup is radiometrically calibrated, we have, from (equations (3.3,3.4)):

$$I_l = \frac{\sin^2(\theta_1 - \theta_2)}{2\sin^2(\theta_1 + \theta_2)} \left(1 + \frac{\cos^2(\theta_1 + \theta_2)}{\cos^2(\theta_1 - \theta_2)}\right) I$$

Let $r = \frac{I_L}{I_{in}}$; we then get the equation

$$r = \frac{\sin^2(\theta_1 - \theta_2)}{2\sin^2(\theta_1 + \theta_2)} \left(1 + \frac{\cos^2(\theta_1 + \theta_2)}{\cos^2(\theta_1 - \theta_2)}\right)$$

Note that although we are considering the case of reflection here, the refracted angle θ_2 nevertheless appears in the equation, due to the "light fall-off" caused by the object's refractive property.

In the following let us denote $s_1 = \sin(\theta_1)$ and $c_1 = \cos(\theta_1)$. Further, we use $\lambda = \frac{n_1}{n_2}$ as the relative refractive index, $D = 1 - \lambda^2 s_1^2$ and $c_2 = \pm\sqrt{1 - s_2^2} = \pm\sqrt{1 - \lambda^2 s_1^2} = \pm\sqrt{D}$. Thus, from the above equation, we get

$$\begin{aligned}
r &= \frac{R_s + R_p}{2} \\
&= \frac{\sin^2(\theta_1 - \theta_2) + \cos^2(\theta_1 + \theta_2)\tan^2(\theta_1 - \theta_2)}{2\sin^2(\theta_1 + \theta_2)} \\
&= \frac{\{s_1 c_2 - c_1 s_2\}^2 + \{c_1 c_2 - s_1 s_2\}^2 \frac{\{t_1 - t_2\}^2}{\{1 + t_1 t_2\}^2}}{2\{s_1 c_2 + c_1 s_2\}^2} \\
&= \frac{\left\{\pm s_1\sqrt{D} - \lambda c_1 s_1\right\}^2 + \left\{\pm c_1\sqrt{D} - \lambda s_1^2\right\}^2 \frac{\left\{t_1 \mp \lambda \frac{s_1}{\sqrt{D}}\right\}^2}{\left\{1 \pm t_1 n \frac{s_1}{\sqrt{D}}\right\}^2}}{2\left\{\pm s_1\sqrt{D} + \lambda c_1 s_1\right\}^2} \\
&= \frac{\left\{\pm s_1\sqrt{D} - \lambda c_1 s_1\right\}^2 + s_1^2\left\{\pm c_1\sqrt{D} - \lambda s_1^2\right\}^2 \frac{\left\{\frac{\sqrt{D} \mp \lambda c_1}{c_1\sqrt{D}}\right\}^2}{\left\{\frac{c_1\sqrt{D} \pm n s_1^2}{c_1\sqrt{D}}\right\}^2}}{2\left\{\pm s_1\sqrt{D} + \lambda c_1 s_1\right\}^2} \\
&= \frac{\left\{\pm s_1\sqrt{D} - \lambda c_1 s_1\right\}^2 + s_1^2\left\{\pm c_1\sqrt{D} - \lambda s_1^2\right\}^2 \left\{\frac{\sqrt{D} \mp \lambda c_1}{c_1\sqrt{D} \pm \lambda s_1^2}\right\}^2}{2\left\{\pm s_1\sqrt{D} + \lambda c_1 s_1\right\}^2} \\
&= \frac{\left\{\pm s_1\sqrt{D} - \lambda c_1 s_1\right\}^2 \left\{c_1\sqrt{D} \pm \lambda s_1^2\right\}^2 + s_1^2\left\{\pm c_1\sqrt{D} - \lambda s_1^2\right\}^2 \left\{\sqrt{D} \mp \lambda c_1\right\}^2}{2\left\{\pm s_1\sqrt{D} + \lambda c_1 s_1\right\}^2 \left\{c_1\sqrt{D} \pm \lambda s_1^2\right\}^2}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\left\{\pm c_1 s_1(1-\lambda^2 s_1^2) \mp \lambda^2 c_1 s_1^3 + \lambda s_1\sqrt{D}(s_1^2-c_1^2)\right\}^2 + s_1^2\left\{\pm c_1(1-\lambda^2 s_1^2) \pm \lambda^2 c_1 s_1^2 - \lambda\sqrt{D}\right\}^2}{2\left\{\pm c_1 s_1(1-\lambda^2 s_1^2) \pm \lambda^2 c_1 s_1^3 + \lambda s_1\sqrt{D}\right\}^2} \\
&= \frac{\left\{\pm c_1 s_1(1-2\lambda^2 s_1^2) + \lambda s_1\sqrt{D}(s_1^2-c_1^2)\right\}^2 + s_1^2\left\{\pm c_1 - \lambda\sqrt{D}\right\}^2}{2\left\{\pm c_1 s_1 + \lambda s_1\sqrt{D}\right\}^2} \\
&= \frac{c_1^2(1-2\lambda^2 s_1^2)^2 + \lambda^2(2s_1^2-1)^2(1-\lambda^2 s_1^2) + c_1^2 + \lambda^2(1-\lambda^2 s_1^2) \pm 2\lambda c_1\sqrt{D}((1-2\lambda^2 s_1^2)(2s_1^2-1)-1)}{2c_1^2 + 2\lambda^2(1-\lambda^2 s_1^2) \pm 4\lambda c_1\sqrt{D}} \\
&= \frac{1 - s_1^2(n^4+4\lambda^2+1) + 4\lambda^2 s_1^4(\lambda^2+1) - 4\lambda^4 s_1^6 \pm 2\lambda c_1\sqrt{D}(-1+s_1^2(\lambda^2+1)-2\lambda^2 s_1^4)}{n^2+1-s_1^2(\lambda^4+1) \pm 2\lambda c_1\sqrt{D}}
\end{aligned}
$$

We put the terms including $\sqrt{D}$ on one side of the equation and the rest on the other:

$$
\begin{aligned}
& \pm 2\lambda c_1\sqrt{D}\left\{r+1-s_1^2(\lambda^2+1)+2\lambda^2 s_1^4\right\} \\
= \;& 1 - s_1^2(\lambda^4+4\lambda^2+1) + 4\lambda^2 s_1^4(\lambda^2+1) - 4\lambda^4 s_1^6 - r\left\{\lambda^2+1-s_1^2(\lambda^4+1)\right\}
\end{aligned}
$$

After squaring both sides of the equations and bringing them back together, we get:

$$
\begin{aligned}
0 = \;& -r^2(\lambda^2-1)^2((\lambda^2+1)s_1^2-1)^2 \\
& + r\left\{8\lambda^4(\lambda^2+1)^2 s_1^8 - 8\lambda^2(\lambda^2+1)(\lambda^4+4\lambda^2+1)s_1^6 + 2(\lambda^8+12\lambda^6+30\lambda^4+12\lambda^2+1)s_1^4 - 2(\lambda^6+14\lambda^4+13\lambda^2+2)s_1^2 + 10\lambda^2+2\right\} \\
& - 4\lambda^4(\lambda^2-1)^2 s_1^8 + 4\lambda^2(n^6-3\lambda^4-\lambda^2+1)s_1^6 - (\lambda^8-4\lambda^6+18\lambda^4-4\lambda^2-1)s_1^4 - 2(5\lambda^4+2\lambda^2-1)s_1^2 + 4\lambda^2 - 1
\end{aligned}
$$

Finally, rearranging the terms with respect to s_1, one obtains the following constraint in θ_1

$$
\begin{aligned}
& s_1^8\left\{4\lambda^4\left(2r(\lambda^2+1)^2-(\lambda^2-1)^2\right)\right\} \\
& + s_1^6\left\{4\lambda^2\left((\lambda^6-3\lambda^4-\lambda^2+1) - 2r(\lambda^2+1)(\lambda^4+4\lambda^2+1)\right)\right\} \\
& + s_1^4\left\{2r(\lambda^8+12\lambda^6+30\lambda^4+12\lambda^2+1) - r^2(\lambda^4-1)^2 - (\lambda^8-4\lambda^6+18\lambda^4-4\lambda^2-1)\right\} \\
& + s_1^2\left\{2r^2(\lambda^2-1)^2(\lambda^2+1) - 2r(\lambda^6+14\lambda^4+13\lambda^2+2) - 2(5\lambda^4+2\lambda^2-1)\right\} \\
& + \left\{2r(5\lambda^2+1) - r^2(\lambda^2-1)^2 + 4\lambda^2 - 1\right\} = 0 \qquad (3.17)
\end{aligned}
$$

Here, $\lambda = \frac{n_1}{n_2}$ and $s_1 = \sin\theta_1$.

We observe that this is a quartic polynomial in $\sin^2\theta_1$, i.e. the computation of θ_1 can be considered as being (close to) a closed-form solution.

Given an estimate for θ_1, the depth d can be computed by triangulating the incident and reflected rays (since camera centre and 3D correspondence are known, and knowledge of θ_1 gives normal and reflected ray directions) Among the up to 8 possible real solutions for θ_1, at most 4 will correspond to a positive depth, i.e. a surface point in front of the camera. Finding a unique θ_1 and thus depth, is made possible in most practical circumstances, as follows. First, the absolute value of θ_1 must be below 90°. Second, θ_1 is typically (much) larger than 30°, due to the practical setup which requires that the camera have both a reflected and a direct view of the monitor. Consider the graph of r as a function of θ_1 for the refractive index of water ($n_2 = 1.33$), in figure (ref figure here) (here, both the camera and the monitor, are in air, i.e. $n_1 = 1$). One observes that for the values of r associated with $\theta_1 \in (30°, 90°)$, there is a unique θ_1 producing these values. To be precise, θ_1 is unique up to sign, but only the positive solution corresponds to a depth/point in front of the camera.

Let us summarize the above findings. From images acquired with a completely static setup, we are able to compute the depth of the transparent object, for each pixel in which a reflection is visible. To do so, we need to know the refractive index of the object's material. A unique solution for depth is possible in a large range of practical conditions. In case these are not fulfilled, one may still use bounds on the object depth to get a unique solution.

3.5.2 Double Bounce

In this section we consider the problem of solving for the depth of surfaces in the presence of two bounces. We first analyse the simplest case, when a rectangular slab separates the camera and the reference. To further simplify the case, we consider the camera plane, the reference planes and the planes of the slab to be parallel. Hence the problem reduces to the estimation of the two "depths" of the slab's planes. Figure 3.3 illustrates this case and the associated coordinate system in which we operate. Note that like opaque objects, each point on the surface of the slab produces one image point. However, unlike opaque objects, the correspondence between image points of a particular surface point in C_1 and C_2 cannot be computed since the surface point does not have an "appearance" but simply transfers a fraction of the appearance of its background. In fact, every back-projected image point from C_1 or C_2 intersects the slab surface at two points and hence has two "depths"

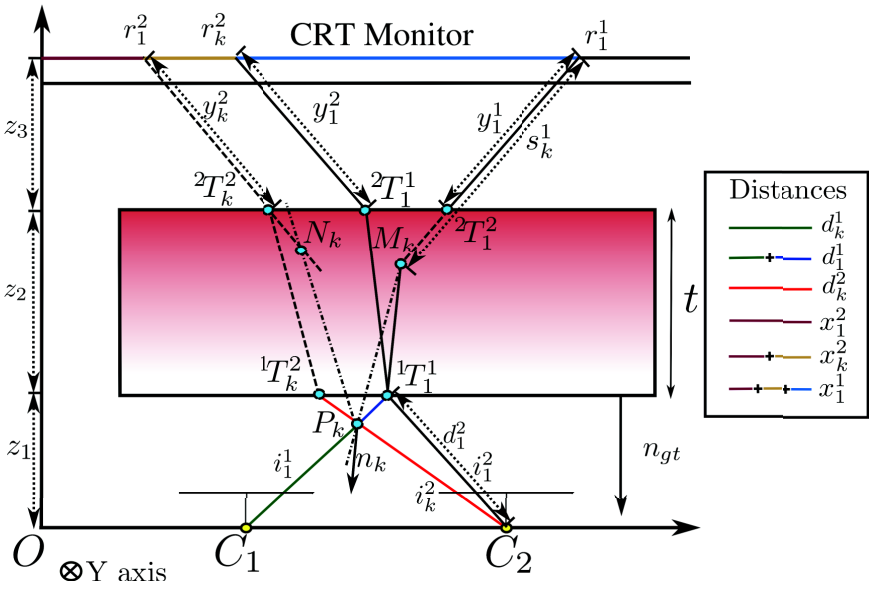


Figure 3.3: Figure representing the simplest double bounce case. Two cameras look at a rectangular glass slab at the CRT monitor placed at two positions, producing overall 4 images. Note that in 3D the intersection point N_k might not actually exist for all values of d_k.

associated with it. Since there are two depth values associated with every ray, and no correspondence information is available between cameras, we need two views to achieve minimal requirements for a solution. We also need two monitor positions because we need to know the final refracted ray's position and direction.

In Figure 3.3, consider a ray with direction i_1^1 from camera C_1. This ray intersects the slab twice before ending up on monitor pixel X_1, with final refracted ray y_1^1. Because of the presence of parallel slabs, we have $i_1^1 = y_1^1$. This light-path also contains two intersection points with the rectangular slab, ${}^1T_1^1$ and ${}^2T_1^1$. In order to reconstruct this light-path, we hypothesize the positions of these intersection points. We then verify the hypothesis using camera C_2. The verification is done in two stages. In the first stage we verify the validity of the light-path produced in C_2 because of the normal and depth values obtained using the first light-path. In the second stage, we verify whether the light-paths satisfy their photometric constraints (using Fresnel equations). Note that although there is a 2D set of values for $({}^1T_1^1, {}^2T_1^1)$, not all values satisfy the Fresnel equations for the first light-path itself.

We model the different light-path hypotheses using two parameters d_k^1 and ${}^1\theta_k^1$. While d_k^1 models the depth of the point ${}^1T_1^1$ from the point C_1 in the direction of i_1^1, ${}^1\theta_k^1$ models the angle between i_1^1 and the normal for the first bounce, n_k. Note that using an angle formulation is possible in this case since we know the plane of refraction Π_1, as it contains C_1, X_1 and i_1^1. Also, generating normal hypotheses in this plane ensures that the point ${}^2T_1^1$ exists. The remaining parameters used in our derivation are expressed in Figure 3.3.

Firstly, we look at the geometric constraint. This involves identifying the conditions for the refracted ray from C_2 to intersect its corresponding ray from X_k along i_k^2. To do this, we need to derive the values of n_k, d_k and the direction of the middle segment of both light-paths. We define n_k and the middle segment m_k^1 as

$$
\begin{aligned}
n_k &= i_1^1 \cos({}^1\theta_k^1) + j_1^1 \sin({}^1\theta_k^1) && (3.18)\\
{}^1\theta_k^{1'} &= \sin^{-1}(\lambda \sin({}^1\theta_k^1)) && (3.19)\\
\bar{m_k^1} &= n_k \cos({}^1\theta_k^{1'}) + \lambda(i_1^1 - n_k \cos({}^1\theta_k^1)) && (3.20)\\
&= i_1^1 \cos({}^1\theta_k^1)\cos({}^1\theta_k^{1'}) + \lambda i_1^1 \sin^2({}^1\theta_k^1) + j_1^1 \sin({}^1\theta_k^1)\cos({}^1\theta_k^{1'}) - \\
&\quad j_1^1 \sin({}^1\theta_k^{1'})\cos({}^1\theta_k^1) && (3.21)\\
&= i_1^1 \cos({}^1\theta_k^1 - {}^1\theta_k^{1'}) + j_1^1 \sin({}^1\theta_k^1 - {}^1\theta_k^{1'}) && (3.22)
\end{aligned}
$$

where $m_k^1 = \frac{\bar{m_k^1}}{\|\bar{m_k^1}\|}$. Now, the first segment of the light-path of C_2, and its angle with normal n_k, ${}^1\theta_k^2$ can be defined as

$$i_k^2 = \frac{C_1 + d_k i_1^1 - C_2}{\|C_1 + d_k^1 i_1^1 - C_2\|} = \frac{\bar{i_k^2}}{\|\bar{i_k^2}\|} \tag{3.23}$$

$${}^1\theta_k^2 = \cos^{-1}(n_k \cdot i_k^2) \tag{3.24}$$

In a similar manner to the approach described for m_k^1, we can derive the equation for m_k^2 as

$$\bar{m_k^2} = n_k \cos({}^1\theta_k^{2'}) + \lambda(i_k^2 - n_k \cos({}^1\theta_k^2)) \tag{3.25}$$

$$= n_k \sqrt{(1 - \lambda^2 \sin^2({}^1\theta_k^2))} + \lambda(i_k^2 - n_k(n_k \cdot i_k^2)) \tag{3.26}$$

$$= n_k \sqrt{(1 - \lambda^2) + \lambda^2 (n_k \cdot i_k^2)^2} + \lambda(i_k^2 - n_k(n_k \cdot i_k^2)) \tag{3.27}$$

Finally, we can define the geometric constraint as the condition that the two rays $X_2 + \alpha \bar{i_k^2}$ and $C_2 + \gamma i_k^2 + \beta \bar{m_k^2}$ intersect (where γ is $\|d_k^1 i_1^1 + C_1 - C_2\|$). In other words

$$C_2 + \gamma i_k^2 + \beta \bar{m_k^2} = X_2 + \alpha \bar{i_k^2} \tag{3.28}$$

$$(C_2 - X_2) \times \bar{i_k^2} + 0 + \beta \bar{m_k^2} \times \bar{i_k^2} = 0 \tag{3.29}$$

$$(C_2 - X_2) \times \bar{i_k^2} = -\beta \bar{m_k^2} \times \bar{i_k^2} \tag{3.30}$$

The above constraint really specifies that for a valid set of parameters the planes given by $(X_2 - C_2), i_k^2$ and n_k, i_k^2 are one and the same. By using equation (3.25), we get

$$
\begin{aligned}
(C_2 - X_2) \times \bar{i_k^2} &= -\beta \left(n_k \times \bar{i_k^2} \cos({}^1\theta_k^{2'}) + \right. \\
&\quad \left. \lambda(0 - n_k \times \bar{i_k^2} \cos({}^1\theta_k^2)) \right) && (3.31) \\
&= -\beta (n_k \times \bar{i_k^2})(\cos({}^1\theta_k^{2'}) - \cos({}^1\theta_k^2)) && (3.32) \\
((C_2 - X_2) \times \bar{i_k^2}) \times (n_k \times \bar{i_k^2}) &= 0 && (3.33)
\end{aligned}
$$

$$
\begin{aligned}
\big(\underbrace{(C_2 - X_2) \times (C_1 - C_2)}_{a} + d_k^1 \underbrace{(C_2 - X_2) \times i_1^1}_{b} \big) & \\
\times \big(\cos({}^1\theta_k^1) \underbrace{i_1^1 \times (C_1 - C_2)}_{x} + & \\
\sin({}^1\theta_k^1) \underbrace{j_1^1 \times (C_1 - C_2)}_{y} + d_k^1 \sin({}^1\theta_k^1) \underbrace{(j_1^1 \times i_1^1)}_{z} \big) &= 0 \quad \text{(using equations (3.23 \& 3.18))} \quad (3.34)
\end{aligned}
$$

Note that vectors a, b, x, y, z are all known quantities. We can further simplify this to remove the effect of ${}^1\theta_k^1$ using the following steps

$$
\begin{aligned}
a \times x \cos({}^1\theta_k^1) + b \times x \, d_k^1 \cos({}^1\theta_k^1) + a \times y \sin({}^1\theta_k^1) + & \\
b \times y \, d_k^1 \sin({}^1\theta_k^1) + a \times z \, d_k^1 \sin({}^1\theta_k^1) + b \times z (d_k^1)^2 \sin({}^1\theta_k^1) &= 0 && (3.35) \\
b \times x \big(a \times x \big(a \times y + b \times y \, d_k^1 + a \times z \, d_k^1 + b \times z (d_k^1)^2 \big) \big) \sin({}^1\theta_k^1) &= 0 && (3.36) \\
b \times x \big(a \times x \big(a \times y + b \times y \, d_k^1 + a \times z \, d_k^1 + b \times z (d_k^1)^2 \big) \big) &= 0 && (3.37)
\end{aligned}
$$

This is a 2nd degree equation in d_k^1 with two solutions. The two solutions can easily be disambiguated by computing the value for ${}^1\theta_k^1$ using equation (3.35) (by solving for $\tan({}^1\theta_k^1)$). Further, observe that since $y_1^1 = i_1^1$ and $y_k^2 = i_k^2$, the angles of the first bounce for the second light-path $({}^1\theta_k^2, {}^1\theta_k^{2'})$ are the same as the angles for the second bounce (in reverse because of reciprocal relative refractive indices).

The photometric constraint computed using Fresnel equations can be given as

$$
\begin{aligned}
I_{rr} &= \frac{1}{2} ({}^1T_s \, {}^2T_s + {}^1T_p \, {}^2T_p) I_{in} && (3.38) \\
r = \frac{I_{rr}}{I_{in}} &= \frac{1}{2} ({}^1T_s \, {}^2T_s + {}^1T_p \, {}^2T_p) && (3.39)
\end{aligned}
$$

Denoting $s_1 = \sin({}^1\theta_k^2), s_2 = \sin({}^1\theta_k^{2'}) = \lambda s_1$ (where $\lambda = \frac{n_1}{n_2}$), and $c_1 = \cos({}^1\theta_k^2)$ and $c_2 = \sqrt{1-s_2^2} = \sqrt{1-\lambda^2 s_1^2} = \sqrt{D}$ (where $D = 1-\lambda^2 s_1^2$), we get

$$
\begin{aligned}
{}^1T_s &= \frac{4\sin^2({}^1\theta_k^{2'})\cos^2({}^1\theta_k^2)}{\sin^2({}^1\theta_k^2 + {}^1\theta_k^{2'})} && (3.40)\\
&= \frac{4\lambda^2 s_1^2 c_1^2}{(s_1c_2 + s_2c_1)^2} && (3.41)\\
&= \frac{4\lambda^2 s_1^2 c_1^2}{(\lambda s_1 c_1 \pm s_1\sqrt{D})^2} = \frac{4\lambda^2 s_1^2 c_1^2}{\lambda^2 s_1^2 c_1^2 + s_1^2 D \pm 2\lambda s_1^2 c_1\sqrt{D}} && (3.42)\\
&= \frac{4\lambda^2 c_1^2}{\lambda^2 c_1^2 + D \pm 2\lambda c_1\sqrt{D}} = \frac{4\lambda^2 c_1^2}{(\lambda c_1 \pm \sqrt{D})^2} && (3.43)\\
{}^2T_s &= \frac{4\sin^2({}^1\theta_k^2)\cos^2({}^1\theta_k^{2'})}{\sin^2({}^1\theta_k^2 + {}^1\theta_k^{2'})} && (3.44)\\
&= \frac{4(1-\lambda^2 s_1^2)}{\lambda^2 c_1^2 + D \pm 2\lambda c_1\sqrt{D}} = \frac{4D}{(\lambda c_1 \pm \sqrt{D})^2} && (3.45)
\end{aligned}
$$

$$
\begin{aligned}
{}^1T_p &= \frac{4\sin^2({}^1\theta_k^{2'})\cos^2({}^1\theta_k^2)}{\sin^2({}^1\theta_k^2 + {}^1\theta_k^{2'})\cos^2({}^1\theta_k^2 - {}^1\theta_k^{2'})} && (3.46)\\
&= \frac{4\lambda^2 c_1^2}{(\lambda c_1 \pm \sqrt{D})^2(c_1c_2 + s_1s_2)^2} && (3.47)\\
&= \frac{4\lambda^2 c_1^2}{(\lambda c_1 \pm \sqrt{D})^2(c_1^2c_2^2 + s_1^2s_2^2 + 2s_1s_2c_1c_2)} && (3.48)\\
&= \frac{4\lambda^2}{(\lambda c_1 \pm \sqrt{D})^2(c_2^2 + t_1^2s_2^2 + 2t_1s_2c_2)} && (3.49)\\
&= \frac{4\lambda^2}{(\lambda c_1 \pm \sqrt{D})^2(D + \lambda^2 t_1^2 s_1^2 \pm 2\lambda t_1 s_1\sqrt{D})} && (3.50)\\
&= \frac{4\lambda^2}{(\lambda c_1 \pm \sqrt{D})^2(\lambda t_1 s_1 \pm \sqrt{D})^2} && (3.51)\\
&= \frac{4\lambda^2 c_1^2}{(\lambda c_1 \pm \sqrt{D})^2(\lambda s_1^2 \pm c_1\sqrt{D})^2} && (3.52)
\end{aligned}
$$

$$
\begin{aligned}
{}^2T_p &= \frac{4\sin^2({}^1\theta_k^2)\cos^2({}^1\theta_k^{2'})}{\sin^2({}^1\theta_k^2+{}^1\theta_k^{2'})\cos^2({}^1\theta_k^2-{}^1\theta_k^{2'})} && (3.53)\\
&= \frac{4c_2^2}{(\lambda c_1\pm\sqrt{D})^2(c_1^2c_2^2+s_1^2s_2^2+2s_1s_2c_1c_2)} && (3.54)\\
&= \frac{4}{(\lambda c_1\pm\sqrt{D})^2(c_1^2+s_1^2t_2^2+2s_1t_2c_2)} && (3.55)\\
&= \frac{4}{(\lambda c_1\pm\sqrt{D})^2(c_1+s_1t_2)^2} = \frac{4}{(\lambda c_1\pm\sqrt{D})^2(c_1\pm\frac{\lambda s_1^2}{\sqrt{D}})^2} && (3.56)\\
&= \frac{4D}{(\lambda c_1\pm\sqrt{D})^2(c_1\sqrt{D}\pm\lambda s_1^2)^2} && (3.57)
\end{aligned}
$$

931 Putting all these terms in equation (3.39) we get

$$
r = \frac{8\lambda^2c_1^2D}{(\lambda c_1\pm\sqrt{D})^4}+\frac{8\lambda^2c_1^2D}{(\lambda c_1\pm\sqrt{D})^4(c_1\sqrt{D}\pm\lambda s_1^2)^4} \quad (3.58)
$$

932 Let us develop each of these terms before combining them into one set of equations

$$
\begin{aligned}
(\lambda c_1\pm\sqrt{D})^4 &= (\lambda^2c_1^2+D\pm2\sqrt{D}\lambda c_1)^2 && (3.59)\\
&= \lambda^4c_1^4+D^2+4\lambda^2Dc_1^2+2\lambda^2c_1^2D\pm4\lambda c_1D\sqrt{D}\pm4\lambda^3c_1^3\sqrt{D} && (3.60)\\
&= \underbrace{\lambda^4c_1^4+D^2+6\lambda^2Dc_1^2}_{A}\pm\sqrt{D}\underbrace{4\lambda c_1(D+\lambda^2c_1^2)}_{B} && (3.61)\\
(c_1\sqrt{D}\pm\lambda s_1^2)^4 &= (c_1^2D+\lambda^2s_1^4\pm2\lambda s_1^2c_1\sqrt{D})^2 && (3.62)\\
&= c_1^4D^2+\lambda^4s_1^8+4\lambda^2s_1^4c_1^2D+2\lambda^2c_1^2s_1^4D\pm4\lambda s_1^2c_1^3D\sqrt{D}\pm4\lambda^3s_1^6c_1\sqrt{D} && (3.63)\\
&= \underbrace{c_1^4D^2+\lambda^4s_1^8+6\lambda^2s_1^4c_1^2D}_{C}\pm\sqrt{D}\underbrace{4\lambda s_1^2c_1(c_1^2D+\lambda^2s_1^4)}_{E} && (3.64)\\
(\lambda c_1\pm\sqrt{D})^4(c_1\sqrt{D}\pm\lambda s_1^2)^4 &= AC+BED\pm\sqrt{D}(AE+BC) && (3.65)
\end{aligned}
$$

933 Substituting equation (3.65, 3.63) in equation 3.58, and cross-multiplying, we get

$$
r(AC+BED\pm\sqrt{D}(AE+BC)) = 8\lambda^2c_1^2D(1+C\pm\sqrt{D}E) \quad (3.66)
$$

We now isolate terms with $\sqrt{D}$ on one side and then square both sides to get

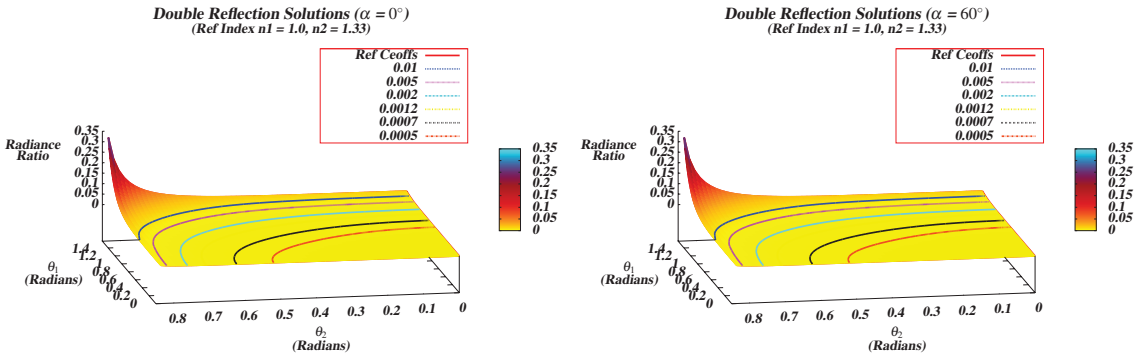
$$\begin{aligned}
\left(\pm\sqrt{D}(rAE + rBC - 8\lambda^2 c_1^2 DE)\right)^2 &= \left(8\lambda^2 c_1^2 D + 8\lambda^2 c_1^2 DC - rAC - rBED\right)^2 && (3.67)\\
D(rAE + rBC - 8\lambda^2 c_1^2 DE)^2 &= (8\lambda^2 c_1^2 D + 8\lambda^2 c_1^2 DC - rAC - rBED)^2 && (3.68)
\end{aligned}$$

In the above equation, the only troublesome term is c_1 since its definition in terms of the variable s_1 involves a square root. However, if all the terms of c_1 are of even powers, this problem does not arise. In order to check this fact, we first expand the different terms in R.H.S and L.H.S.

$$\begin{aligned}
AE &= \left(\lambda^4 c_1^4 + D^2 + 6\lambda^2 D c_1^2\right)\left(4\lambda s_1^2 c_1 (c_1^2 D + \lambda^2 s_1^4)\right) && (3.69)\\
&= 4\lambda s_1^2 c_1 \left(\lambda^4 c_1^4 + D^2 + 6\lambda^2 D c_1^2\right)\left(c_1^2 D + \lambda^2 s_1^4\right) && (3.70)\\
BC &= \left(4\lambda c_1 (D + \lambda^2 c_1^2)\right)\left(c_1^4 D^2 + \lambda^4 s_1^8 + 6\lambda^2 s_1^4 c_1^2 D\right) && (3.71)\\
&= 4\lambda c_1 \left(D + \lambda^2 c_1^2\right)\left(c_1^4 D^2 + \lambda^4 s_1^8 + 6\lambda^2 s_1^4 c_1^2 D\right) && (3.72)\\
8\lambda^2 c_1^2 DE &= 32\lambda^3 s_1^2 c_1^3 \left(c_1^2 D + \lambda^2 s_1^4\right) && (3.73)\\
AC &= \left(\lambda^4 c_1^4 + D^2 + 6\lambda^2 D c_1^2\right)\left(c_1^4 D^2 + \lambda^4 s_1^8 + 6\lambda^2 s_1^4 c_1^2 D\right) && (3.74)\\
BE &= \left(4\lambda c_1 (D + \lambda^2 c_1^2)\right)\left(4\lambda s_1^2 c_1 (c_1^2 D + \lambda^2 s_1^4)\right) && (3.75)\\
&= 16\lambda^2 s_1^2 c_1^2 \left(D + \lambda^2 c_1^2\right)\left(c_1^2 D + \lambda^2 s_1^4\right) && (3.76)\\
8\lambda^2 c_1^2 DC &= 8\lambda^2 c_1^2 D\left(c_1^4 D^2 + \lambda^4 s_1^8 + 6\lambda^2 s_1^4 c_1^2 D\right) && (3.77)
\end{aligned}$$

Observe that the R.H.S terms (equations (3.70, 3.72, 3.73)) all have odd powers of c_1 and the L.H.S terms (equations (3.74, 3.76, 3.77)) all have even powers of c_1. Thus squaring each side would produce terms that are all even powers of c_1. Also, observe that the term A is of degree 4 in s_1, B is of degree 3, C of 8, D of 2 and E of 7. So both the R. H. S and L. H. S in the equation (3.68) are of degree 24.Thus we get a 24 degree equation in s_1 and degree 12 in s_1^2. Thus we get 12 solutions for s_1 since we only consider positive roots.

Note also, that many of the variables involved in the general case of double bounce are predetermined here because of the simplicity of the case. For example, the condition $y_1^1 = i_1^1$ is not true in general. Also, we assume that the entire light-path lies on a single plane, which is not true in the general case. Thus, it is easier to show that a unique solution may be obtained empirically. Figure 3.4 simulates the value of radiance ratio for the space spanned by the two incident angles



(a) (b)

Figure 3.4: Radiance ratio values plotted for various pairs of incident angles. There is no appreciable difference because of a non-zero angle between the two planes of refraction in the double bounce case. Isocontours for various values of radiance ratio are plotted. Note that they are all curves. Note also that their intersection with the curve corresponding to $\theta_2 = \theta_1'$ (case of parallel planes) is a single point.

along a light path. It shows that given refractive index, the values of incident angles that correspond to a particular radiance ratio lie on a curve in this space. A similar case occurs for refractive double bounce.

3.5.3 Practical considerations

The above theory shows that the irradiance of a final light ray in a light-path contains information that could be used to reconstruct the entire light-path. In this section we describe important elements of our experiments to collect radiance measurements for reconstruction. Our experiments consist of three parts. 1) We use an illuminant with known geometry to emit unpolarized light in a desired set of directions. 2) Light from the illuminant interacts with the transparent object, and reflects / refracts off its surface towards the camera after one or two bounces. 3) The camera then captures both the direction and radiance of some reflected / refracted light rays, which is used for reconstruction.

We use Canon and Nikon DSLRs for our experiments. Since we need to capture the position and radiance of an individual light ray, we adopt the pin-hole model for the camera (smallest aperture). Although using the smallest aperture does not guarantee a single ray for every direction, the cone of rays captured by the camera at each CCD pixel can be approximated by a thin cylinder if the

focal length is large, or the objects are far away. As for the focal length, there is a tradeoff between large depth-of-field and imaging the pixels of the monitor. We would like to minimize blur because it confuses radiance measurements, but on the other hand we would like to capture each pixel of the monitor on 1 or more pixels on the camera CCD (this leads to a simpler illumination model for the monitor pixel). We arrived on an acceptable range of focal lengths by trial and error. Finally, for each captured ray, we compute the corresponding pixel on the monitor from which the ray originated using standard methods [5].

Unpolarized illuminant In our experiments, we use a flat CRT monitor (LCD montiors emit polarized light), whose pose is computed with respect to an internally calibrated camera [129]. This is motivated by our need to measure the illumination and the geometry of our illuminant. We capture the light emitted by each pixel of the monitor in several directions in 3D, and fit a smooth model to this data in order to accurately measure radiance of the illumination incident on the transparent objects.

Interreflections A common problem with measuring illumination reflected / refracted off specular transparent objects is interreflections. They not only corrupt the radiance measurement, but also pose a problem to correspondence estimation. In order to remove the effect of interreflections, we use the algorithm of Nayar *et al.* [99]. Instead of using a projector to light the scene, we use the CRT monitor. We project low frequency checkerboard patterns that are shifted cyclically, and use it to compute the direct and global components of the scene. Figure 3.10 shows a result for one typical scenario.

Calibration We internally calibrate the camera in order to compute the direction of the captured rays. We also externally calibrate the monitor w.r.t the camera. We use the monitor as the reference frame of the coordinate system, which is beneficial in the case of multiple cameras. For irradiance measurement, we first extract an unprocessed image from the RAW files using dcraw [1]. We then remove the global component of this image, and then extract direct irradiance measurements. Note that all images shown in the results section are extracted from RAW files.

[1] Command : `dcraw -r 1 1 1 1 -D 0 -H 1 -q 3 -4 -v` for Canon cameras. We drop the `-D` option for Nikon cameras.

Correspondence Acquiring correspondence between pixels on the monitor and pixels on the camera that correspond to the same light-path becomes slightly cumbersome when transparent objects are involved [4]. In order to overcome this, we first remove global components from images used for correspondence by applying the theory of [99] to these binary images, and then use a graph based approach to enforce spatial smoothness in the image while extracting correspondences. We use gray codes for robustness to intensity measurement errors, and in our experience the global component removed images are easier to process for correspondences than images which contain both direct and global components.

3.5.4 Estimating CRT Illumination Model

In this section, we give a slightly more detailed description of how we calibrate the illumination of the CRT monitor that we use in our experiments. This is essential since without an accurate calibration, we will not obtain accurate ratios of radiances. Throughout this section we assume a flat CRT monitor screen. Our experiments use a flat CRT monitor screen with a resolution of 768×1024 pixels.

Consider a camera viewing the monitor screen, with no object in between (Figure 3.5). Also consider that the pose of the camera w.r.t the monitor plane is known, and that several such images have been captured with the camera at different positions w.r.t the monitor (with pose known in each case, in a common coordinate system).

Consider the j^{th} pixel on the monitor screen. Let it have a maximum radiance of I_j, captured when the camera looks directly at the screen (camera plane parallel to monitor plane). In other poses, when the camera captures its image, this pixel contributes to 1 or several pixels on the CCD. Also, its contribution to each CCD pixel might be partial (with other monitor pixels also contributing to the same CCD pixel) or whole (no other contributions to that CCD pixel). Thus, a CCD pixel receives a percentage $H(j)$ of a monitor pixel's radiance. Note that in this aspect of the model, we deviate from the pin-hole model towards a model where each ray is represented as a cylinder, as shown in Figure 3.5.

Also, when a camera CCD pixel imaging part of the monitor is back-projected into world coordinates, the back-projected ray lands on the monitor screen at an angle w.r.t to the normal of the monitor plane. Let this angle be called θ_j for each monitor pixel (actually more than one CCD pixels might back-project onto the j^{th} pixel, but since all these pixels will be in a small neighborhood, it is

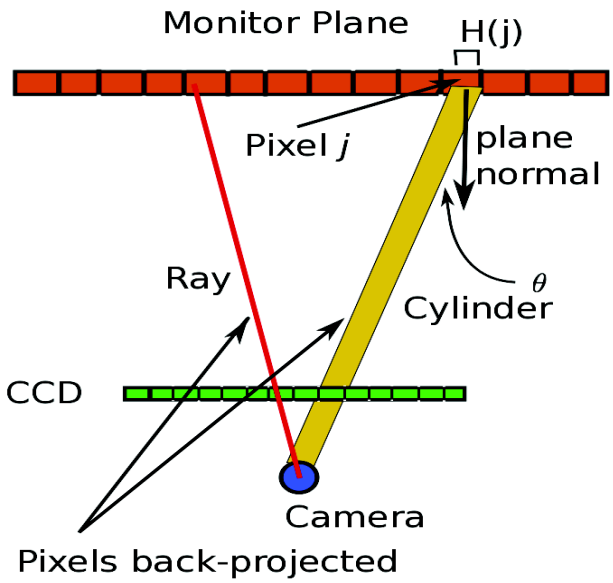


Figure 3.5: Illustration depicting various elements of the illumination model. Two camera pixels are back-projected, while a pin-hole model (used for geometric calibration) assumes a back-projected ray, the illumination model assumes a back-projected cylinder. This cylinder strikes pixel j on the monitor and captures $H(j)$ percent of its illumination in the direction θ.

reasonable to assume they will make the same angle with the monitor normal).

Finally, if the pixel j has a finite area A_j, then the radiance captured by the back-projected ray and hence the camera is proportional to $\cos(\theta_j)$ because of the change in the solid angle subtended by the pixel.

Putting all this together, we can explain the radiance observed at camera CCD pixel l in pose k as

$$I_l = \sum_{m=1}^{N_j} I_m \cos(\theta_m^k) H(m^k) \rho(\theta_m^k) \tag{3.78}$$

where N_j is the bunch of pixels that project onto pixel l. Finally, if we assume I_m and θ_m to be constant over a small neighborhood and we assume the camera zooms in on the monitor pixels so that 1 monitor pixel projects onto several CCD pixels, we get.

$$I_l = H(j^k) I_j \cos(\theta_j^k) \rho(\theta_j^k) \tag{3.79}$$

In the above equations, all the variables are known except $\rho(\theta_j^k)$ which explains the varying amount of radiance emitted by the j^{th} monitor pixel at angle θ_j^k w.r.t to the monitor plane normal. Note that by construction, this function is assumed to be symmetric about the normal.

The value of $H(j^k)$ is computed using the homography between the image of the monitor and the monitor pixels themselves (which is known by calibration). For each CCD pixel, we generate several equally spaced points within the unit area of the pixel, and transfer them to monitor pixel coordinates using the homography. We determine the value of $H(j^k)$ based on how many of the generated points fall within the boundaries of the monitor pixel j. θ_j^k is computed trivially using the camera pose w.r.t monitor plane.

Figure 3.6 shows the variation of ρ with θ for 6 different pixels on the monitor. Notice how the function has a quadratic fall-off with angle. We fit a quadratic curve to this function and use its parameters in our depth estimation algorithm.

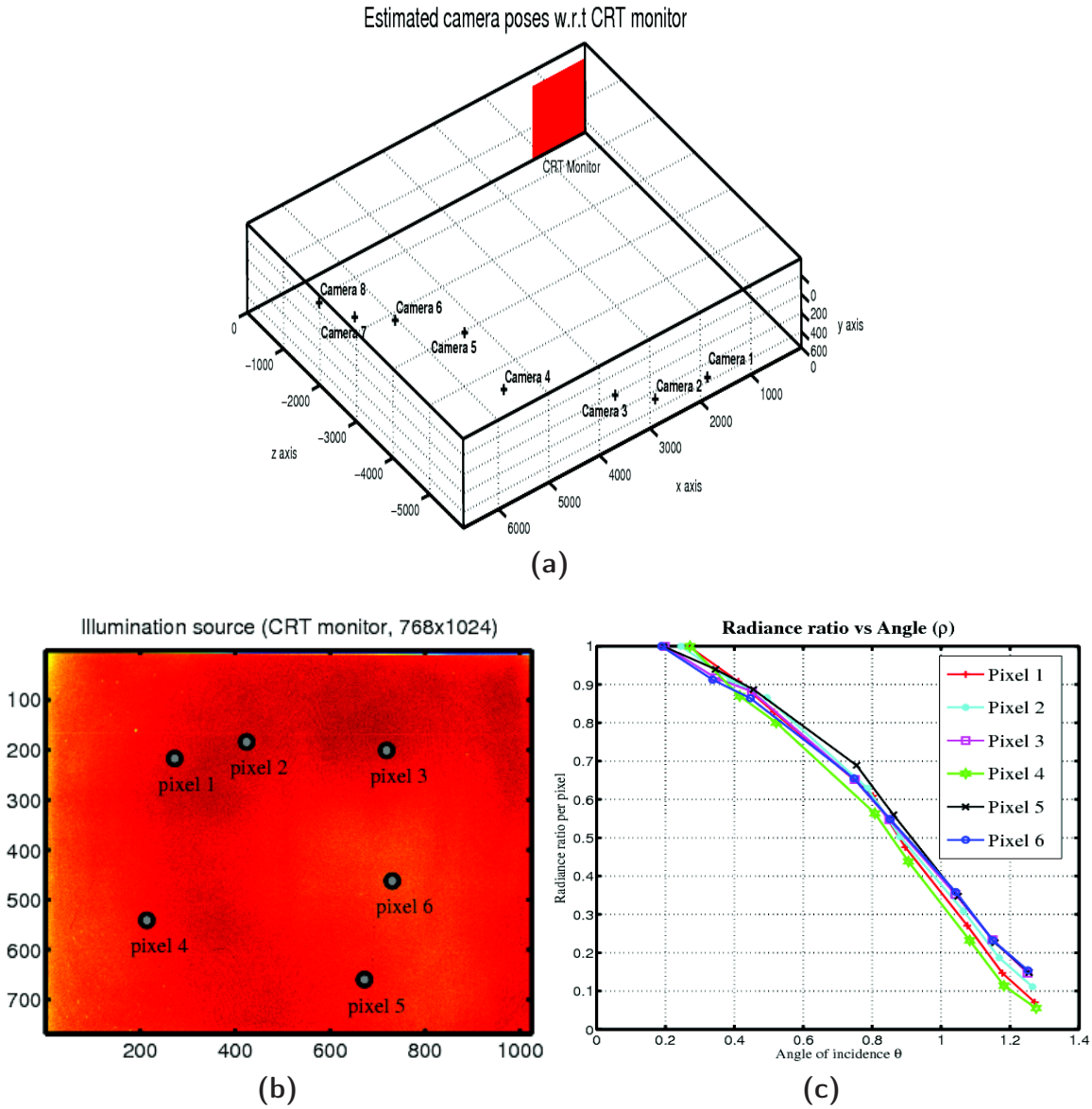


Figure 3.6: Illumination calibration experiment (a) 8 Camera poses w.r.t. CRT monitor plane, used to measure ρ (b) 6 Pixels on the monitor for which ρ values vs θ are plotted in (c). Observe the quadratic nature of ρ.

Figure 3.7: Images of datasets included in this supplementary material, and some images of their acquisition setup

3.6 Experiments and Results

In the previous sections, we showed that radiance ratios could be used to reconstruct transparent surfaces, which can help in reducing the number of measurements required for reconstruction. We also list some practical aspects that are relevant to acquiring image data. In this section, we show results of three experiments. The first two experiments demonstrate the accuracy of 3D reconstruction using our method. The third and fourth experiment show our results on extremely complex scenarios, and we show that just the application of what we theorize above can give surprisingly good results.

Experiment 1: Synthetic dataset Figure 3.9 shows various results for single bounce reconstruction of a sample sinusoidal object when the 3D correspondence is noisy. We compare with two light-path triangulation approaches, one in which the 3D correspondences along a light-path are close to each other and the object (LP-1), and one in which they are far (LP-2). In both cases, noise is added to the farthest 3D correspondence. While one case (LP-1) is sensitive to noise, the second case (LP-2) is robust but impractical. Our approach however, gives a reliable normal map even if the depth is slightly perturbed (compared to LP-2). Note that noise percentage is calculated as ratio of distance between noisy and ground truth data and distance of ground truth and object. Other results w.r.t camera noise and refractive index mismatch are present in the supplementary materials.

Experiment 2: "Water Sequence" Figure 3.10 (Left column) shows some images acquired in order to reconstruct the surface of water in a plastic bowl. The bowl is around 10 cms in diameter, and is placed about 2.5 meters from the camera. This is a scene with a very simple 3D structure (a plane) and given the smooth surface of the bowl, it also has minimal (but not negligible) interreflections and caustics.

Because of the planar nature of the scene, we compute correspondence by simply computing a homography between the reflected image and the direct image of a photograph displayed on the monitor. This homography, adjusted for the internal calibration of the camera and the aspect ratio of the pixels on the monitor, can now be used to compute the normal of the plane [129].

After computing the direct image, we use the homography to compute the ratio of directly observed and reflected radiances. Finally, we hypothesize and test individual pixels for various values of depth, and record the value that best fits the radiance ratio. Figure 3.10 shows our result (the ripples in

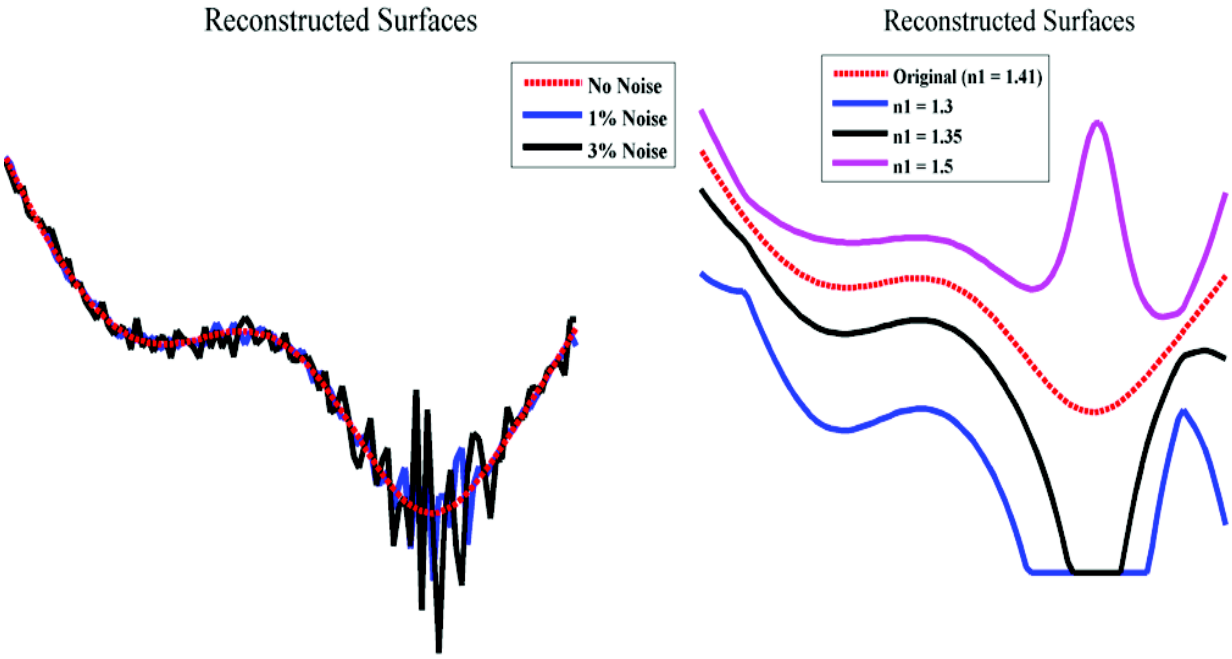


Figure 3.8: Simulation results for photometric error and refractive index mis. We simulated a curved object, and captured radiometric information from a camera according to Fresnel theories. We then added noise to this data (left), or used a slightly different refractive index (right) to reconstruct the surface. Notice that for noise, although the noise was added to the radiometric information that was recorded, the radiance ratio is considerable stable w.r.t the noise (since the denominator in the radiance ratio, which is the illuminant, has a high value). This results in the reconstructions being reasonably stable. However, when the refractive index is changed, depending on the angle of incidence and the depth-normal relationship, the reconstruction accuracy changes. While here we list the worst possible scenario, when the camera is close to the object, moving the camera further away like in the case of our datasets considerably strengthens the robustness of our results. Notice also, that in case we underestimate the refractive index the depth variation of the surface remains more or less intact, while its mean distance from camera increases. Thus, even with an underestimate it is possible to get reasonable results.

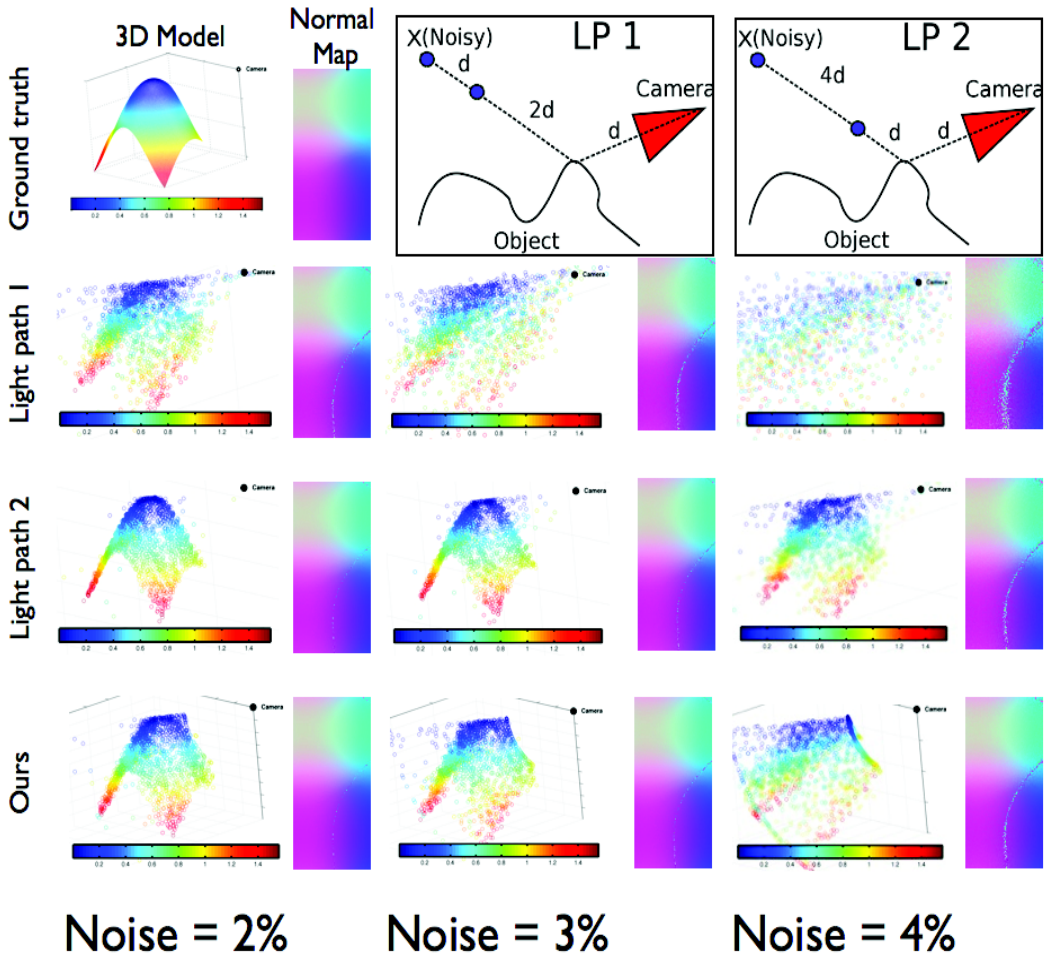


Figure 3.9: Comparison of simulations between our approach and [73]. In LP-1, corresponding 3D points are normally close to the object, which results in increased error in triangulation and normal estimation. Note that in the same scenario, we have much better normal information because of photometric information. While LP-2 is robust because correspondences are far away, its highly impractical since use of LCD's for correspondence is problematic (because of light fall-off, scattering etc.). Details in text.

the reconstruction are just quantized depth values). An alternate result was one obtained by using the algorithm of [129], which uses the knowledge that the object is planar, while our approach estimates per pixel depth. Comparison of the results gives us a mean squared error of around 0.1 cms (we omit correspondence errors while computing this measure), which shows the accuracy of our reconstruction.

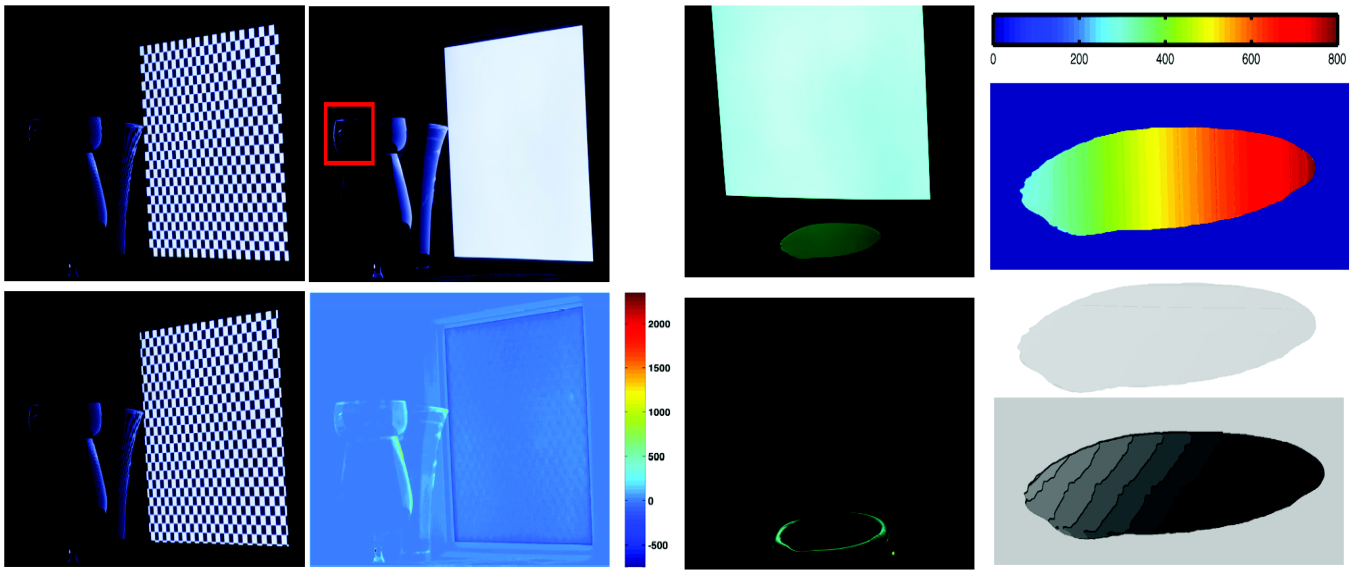


Figure 3.10: (Left Column) Two of 25 images used to compute the direct and global images [99] to remove the effect of interreflections and caustics on radiance measurement. (Middle Left Column) Direct and Global (scaled) components. (Middle Right Column) Difference between the "direct" component and an image taken with a white pattern shown on the monitor. Bottom image is the difference. (Right Column) Correspondence map, Depth map and Reconstructed mesh of "Water" sequence.

Experiment 3: "Wine Glass Sequence" This sequence is very challenging for approaches that use projected texture for reconstruction because of the large interreflections and caustics present in the scene. Note how global components of the image are present even in places where there is no direct light (Figure 3.10, red square). For approaches that typically only use geometric methods [73; 91], the subsurface scattering might throw off correspondence measurements. This can heavily influence reconstruction accuracy. On the other hand, robust measurement of the position and direction of light incident on the glasses from the monitor requires a large set of images to be captured while moving the monitor over, say, an optical bench. In our approach, however, both these errors are avoided because we use a *single* measurement per pixel for reconstruction, and use direct components of the images we capture. Again, like in the case of the "Water Sequence" we hypothesize various depth values along each back-projected pixel, and test their validity using computed radiance ratios. Figure 3.11 shows depth and normal maps computed using our approach. Notice again how we obtain smooth maps even though no smoothness constraints are imposed on the results.

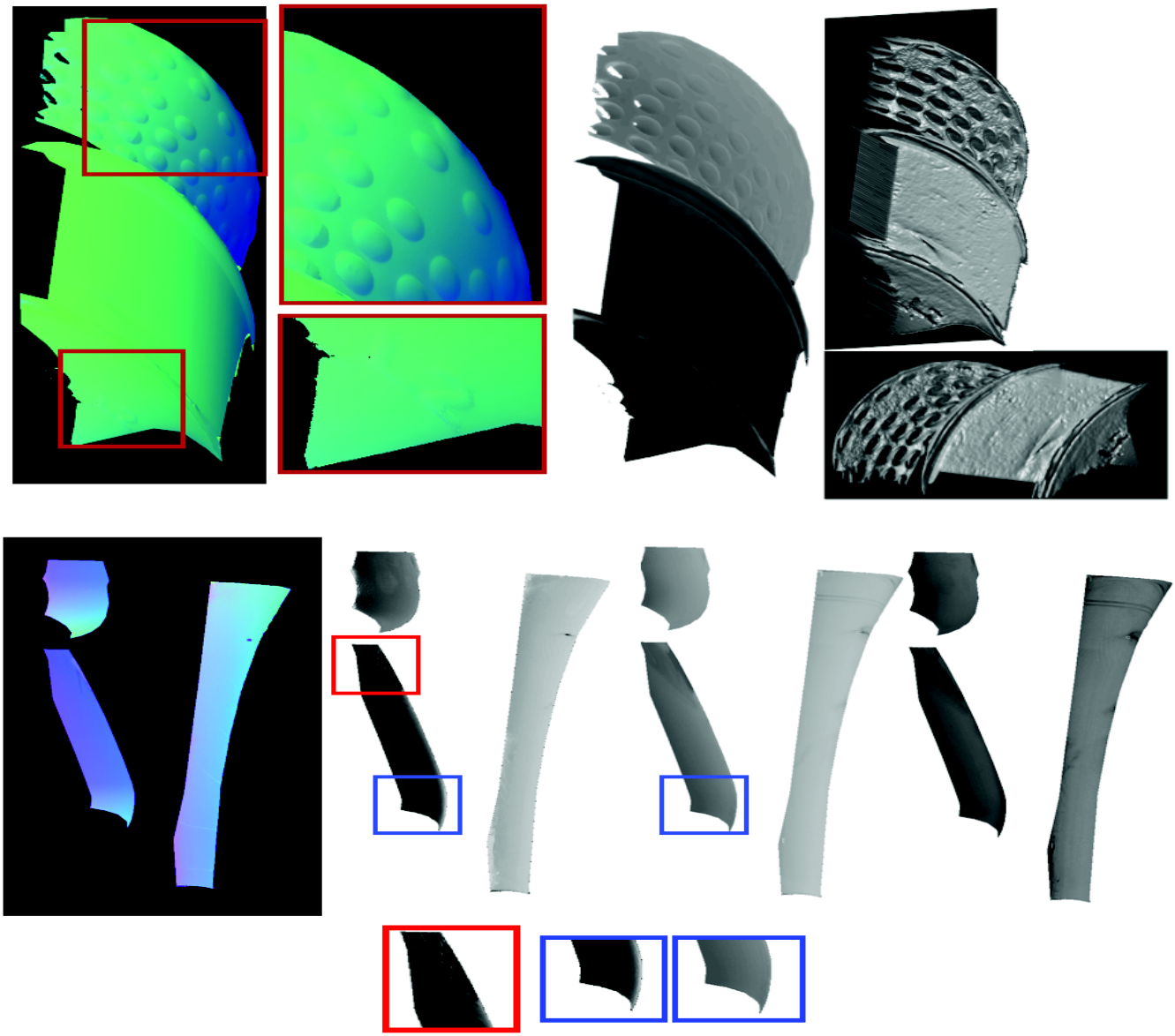


Figure 3.11: (Left top) Normal map of "Fanta bottle" sequence. (insets) Note the fine details captured as a result of radiance ratios. (Centre top) Depth map. Blacker colors are closer to camera. (Right top) Two views of the 3D reconstruction, with lighting to highlight shape variations. Phenomenon like scratches on the bottle, inhomegenous refractive index, violation of single bounce through occlusion are some bad effects, but still reliable reconstructions are achieved. Note that since camera is placed far from the object and monitor, large changes in depth cause small changes in angle. This explains some of the "rough"-ness of the reconstruction. Note also that no smoothing or. optimization is applied for this reconstruction. (Left bottom) Normal map of "Wine glass" sequence. (Left-centre bottom) Depth map. (Right-centre bottom) Depth map produced if no interreflection removal is performed. Notice the lack of depth variation in one of the glasses (compared using insets with blue borders). Some frequency artifacts can be seen (red inset) due to interreflection removal. (Right bottom) Depth difference between the two cases. Best viewed in color.

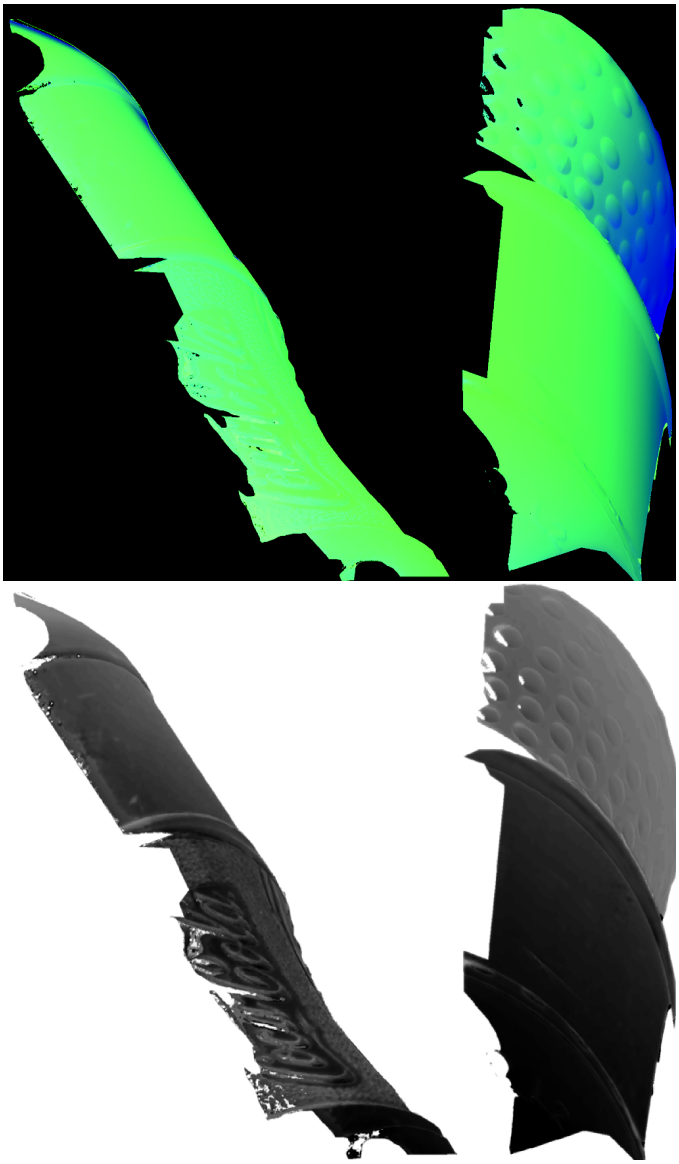


Figure 3.12: Normal map and Depth maps for the cokebottle and fanta sequences.

Experiment 4: "Fanta Bottle Sequence" This sequence highlights the ability of our approach to capture details of a surface. The scene captured consists of a Fanta bottle filled with water. The absence of any large interreflections in this scene results in a very detailed reconstruction as shown in Figure 3.11. Note that optimizing depth and normal simultaneously would serve to remove the artefacts seen in the figure, especially enforcing the depth-normal consistency (differentiation of depth gives normal).

3.6.1 Potential Applications

In the previous section, we showed how using radiance measurements we can extract the shape of transparent objects with minimal measurements even in challenging scenarios. Now we apply the same theory inversely to the problem of calibrating the radiometric response of the camera.

Radiometric calibration is the process of extracting the function that maps image intensities to physical responses of the CCD pixels to incoming light. In short, it maps intensities to irradiances. Earlier methods on radiometric calibration [87] focused on obtaining multiple images at varying exposures in order to obtain *ratios* of intensities that are a function of the ratio of exposure times and the radiometric calibration. Given ratio of exposure times and intensities, one could fit a non-linear model to intensities that would make their ratios converge to exposure ratios [87]. However, this requires several images to compute the ratios, and is limited since camera exposure times only have discrete levels of increments.

However, we could use an image from a scene like the "Water Sequence" where the object is known to be planar, in order to obtain radiance ratios that are related to the depth of the water surface. Such a scene consists of minimal interreflections and scattering, and so intensity ratios from radiometrically uncalibrated images are a function of the nonlinear radiometric responce of the camera only. Since we can compute the depth of the scene using purely geometric approaches [129], we can also obtain a set of "desired" intensity ratios. Moreover, since depth varies continuously, so do the intensity ratios and thus we get *more* data from a single image than from *multiple* images with different exposure. This could potentially be used to radiometrically calibrate the camera.

3.7 Discussion and Conclusion

Reconstruction of transparent objects remains a challenging problem because of the lack of cues that are normally available for other objects. In this paper, we showed the existence of an approach that combines two of the more widely available cues, namely geometric and photometric cues. We showed how this leads to simplified acquisition, a decently robust algorithm, new minimal solutions, and presented challenging scenarios where our approach yielded accurate reconstructions. Applications of our approach lie in valdiating outputs of light manipulation/inference tasks like light-transport matrix estimation and interreflection removal apart from reconstruction and radiometric calibration of cameras.

4

Reconstructing Planar Light-Paths

4.1 Introduction

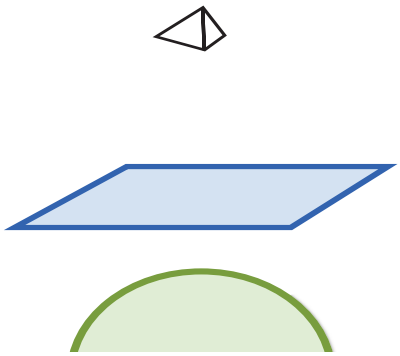
In the second chapter, we considered the case of a camera looking at a scene in a different medium through a planar interface. A generalization of this scenario would be a camera observing a scene through multiple refractive planes (e.g. underwater imaging), which results in distortions and gives the illusion of the scene being closer and magnified. While 3D reconstruction in such scenarios has been analyzed in multi-media photogrammetry [62; 114; 125], such imaging setups have been relatively unaddressed in the computer vision community until recently. Calibrating such a system with multiple layers with unknown layer orientation, distances and refractive indices remains an open and challenging problem.

The fact that such systems do not correspond to a single viewpoint system is known (see, for example, [141]). However, we show that the underlying geometry of rays in such systems actually corresponds to an axial camera. This realization, which has been missing from previous works to the best of our knowledge, allows us to handle multiple layers in a unified way and results in practical and robust algorithms. In fact, we generalize one step further and look at the category of systems with axial light-paths. We observe that in axial systems, light-paths are planar in nature. This observation allows us to collect both mirror and transparent object based systems into a common class of axial systems with planar light-paths. In such systems, we show that it is possible to determine the unknown orientation of the axis independently of the number of layers of refraction or the mirror structure or the refractive indices.

Contents

This results in considerable simplification of the calibration problem via a two-step process, where the axis is computed first. Without such a simplification, calibration is difficult to achieve. Secondly, we show that axis estimation can be mapped to the classical relative orientation problem (essential matrix estimation) for which excellent solutions (e.g. 5-point algorithm [103]) already exist. In fact, calibration can be done using a single plane similar to [137]. We then discuss several cases in which the computed axis estimate might be used to recover the other uknown parameters of each light-path. Our primary contributions are as follows.

- We show that the geometry of rays in flat refraction systems corresponds to an axial camera, leading to a unified theory for calibrating such systems with multiple layers.
- By demonstrating the equivalence with classical essential matrix estimation, we propose efficient and robust algorithms for calibration using planar as well as nonplanar objects.
- We extend the same theory for flat refraction to the case of flat and spherical reflection. In fact we show that some of these cases are analogous.
- We derive theoretical upper bounds of the approximation involved when these models do not fit the actual data, in the case of normal estimation.



4.1.1 Related Work

Maas [83] considered a three layer system assuming that the image plane is parallel to the refractive interfaces. His approach corrects for the radial shift of the projected 3D points using optimization. Treibitz et al. [141] consider a single refraction with known refractive index in an underwater imaging scenario. They assume the distance of the interface as the single unknown parameter (when the camera is internally calibrated) and perform calibration using known depth of a planar checkerboard. The image plane is parallel to the interface in their setup as well. In contrast, we (a) do not assume that the refractive interfaces are frontoparallel, (b) handle multiple layers with unknown layer distances, (c) consider known/unknown refractive indices, and (d) do not assume known pose of the calibrating object. We only assume that the camera is internally calibrated.

3D reconstruction under reflections/refractions has been explored in [22; 26; 73; 92; 124] either for reconstructing the scene or the medium itself. Chen et al. [22] captured two images, with and without a thick glass slab for 3D reconstruction. Both images are required to estimate the orientation of the slab and an additional image to obtain the refractive index. We show that a single set of 2D-3D correspondences from a single photo allows estimating medium thickness as well as refractive index. Other works assume known vertical direction [26] or require several images for calibration [92]. Steger and Kutulakos [73] showed that light-path triangulation becomes degenerate when the entire light-path lies on a plane, which is the case here. Their goal is to compute the shape of the refractive medium, and they consider each light-path independently. In contrast, we have partial knowledge of shape, and light-paths can be parameterized. Thus, we can use information from multiple light-paths to obtain the refractive index. For two refractions (air-medium-air), our analysis is consistent with [73] in that the distance to the medium cannot be estimated. However, we show theoretically that if all refractive indices are different, light paths are not degenerate for any number of layers.

Non-Central/Axial Cameras Pless [109] proposed algorithms for relative motion estimation for calibrated noncentral cameras. Chari and Sturm [23] showed the existence of geometric entities such as fundamental matrix considering refraction. Generic camera calibration algorithms [131] have been proposed to calibrate non-central cameras. Li et al. [77] analyzed the degeneracies in axial cameras for motion estimation. Ramalingam et al. [111] proposed a general framework for calibrating axial cameras using three checkerboards. Their parameterization considers a general axial system and does

not consider the explicit case of a pin-hole camera observing refractive/reflective surfaces, while also involving two rotations/translations. In contrast, we consider the specific case of an axial system passing through the camera center, allowing calibration from a single plane. Models based on radial distortion for calibration assume known center of distortion [139] or model each distortion circle separately [137]. We use a global model with fewer parameters for flat refractive systems.

4.2 Plane of Reflection / Refraction Constraint

Axial Camera: An axial camera is defined as a camera that captures an axial system of rays with the axis passing through the camera center. While the axial nature of many catadioptric systems is already well known [113], we show that an N-layer flat refraction system, with $\mathbf{n}$ being the normal of the flat layer(s), corresponds to an axial camera. The axis is defined as the line parallel to $\mathbf{n}$ passing through the camera center (origin) and let $\hat{\mathbf{a}}$ be its direction vector. Let Π be the plane of refraction (POR) containing the axis and a given camera ray. The normal $\mathbf{n}$ lies on Π. From Snell's law, the incoming ray, the normal and the refracted ray lie on the same plane at the point of refraction. Since the refracted ray from one layer is the incident ray for the next layer (with the same normal), hence, by induction, the entire light-path should lie on Π and all the refracted rays should intersect the axis. Thus, all outgoing rays intersect the axis and the system is axial. Suppose we consider one such light-path, with its first ray direction being denoted by $\mathbf{l}$. Let this light-path eventually pass through the 3D point $\mathsf{R}\mathsf{X}+\mathbf{t}$ (where X is known to us R and $\mathbf{t}$ are not). This 3D point $\mathsf{R}\mathsf{X}+\mathbf{t}$ should also lie on Π. Thus, the plane of refraction constraint for each 3D point can be written as

$$\text{POR Constraint}: \qquad (\mathsf{R}\mathsf{X}+\mathbf{t})^{\top}(\hat{\mathbf{a}}\times\mathbf{l})=0, \tag{4.1}$$

where $(\hat{\mathbf{a}}\times\mathbf{l})$ is the normal to POR. Note that the POR constraint is *independent* of the number of layers N, their thicknesses d_k, and the refractive indices μ_k. It only depends on the axis and pose parameters. It is also independent of the nature of specularity, *i.e.* it applies to both reflection and refraction. Also, note that this *independence* means that light-paths with *different* numbers of bounces could be clubbed together in the POR constraint in order to estimate the axis and pose parameters. This fact is illustrated in Figure 4.2.

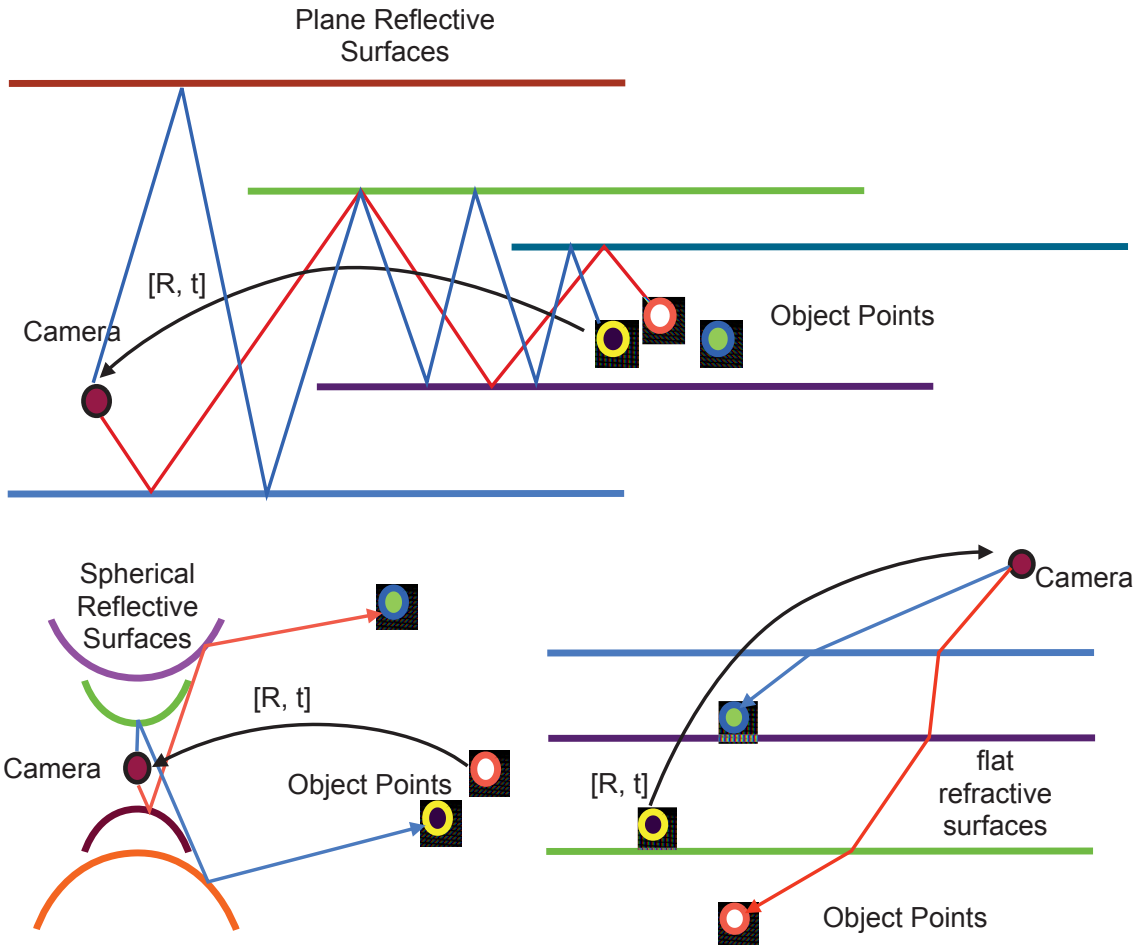


Figure 4.2: Figure showing two light-paths for the plane reflective, the spherical reflective and the flat refractive case. The red and blue light-paths bounce off different surfaces and different numbers of times. However, they can all be clubbed into the same POR constraint to estimate R and $\mathbf{t}_{A^\perp}$.

4.2.1 Axis Computation

Let M(:) be the vector formed by stacking the colums of a matrix M and let $\otimes$ denote the kronecker product. Let $[\hat{\mathbf{a}}]_\times$ be the 3×3 skew-symmetric matrix obtained from 3-vector $\hat{\mathbf{a}}$. The POR constraint can be re-written as

$$\mathbf{l}^\top(\hat{\mathbf{a}} \times (\mathsf{R}\mathbf{X} + \mathbf{t})) = \mathbf{l}^\top \mathsf{E}\mathbf{X} + \mathbf{l}^\top \mathbf{s} = 0, \tag{4.2}$$

where $\mathsf{E} = [\hat{\mathbf{a}}]_\times \mathsf{R}$ and $\mathbf{s} = \hat{\mathbf{a}} \times \mathbf{t}$. Note that $\mathbf{s}^\top \hat{\mathbf{a}} = 0$ and thus the full translation $\mathbf{t}$ cannot be estimated using POR constraints. The component of $\mathbf{t}$ in the direction of the axis, $\mathbf{t}_A$, vanishes in $\mathbf{s}$. Thus, we have 7 degrees of freedom that can be recovered: 2 for the axis, 3 for rotation and 2 for translation.

11-point Linear Algorithm : Stacking equations for 11 $(\mathsf{X}_i, \mathbf{l}_i)$ correspondences, we get a linear system

$$\underbrace{\begin{bmatrix} (\mathsf{X}^{1\top} \otimes \mathbf{l}^{1\top}) & \mathbf{l}^{1\top} \\ \vdots & \vdots \\ (\mathsf{X}^{11\top} \otimes \mathbf{l}^{11\top}) & \mathbf{l}^{11\top} \end{bmatrix}}_{\mathsf{B}} \begin{bmatrix} \mathsf{E}(:) \\ \mathbf{s} \end{bmatrix} = 0, \tag{4.3}$$

where B is an 11×12 matrix whose rank in general is 11. Let $\mathsf{B} = \mathsf{U}\Sigma\mathsf{V}^\top$ be the SVD of B. The solution is given by the right null singular vector of B (last column of V). The scale factor is obtained by setting the norm of E to one.

8-point Algorithm : Notice the striking similarity between our E matrix ($[\hat{\mathbf{a}}]_\times \mathsf{R}$) and the essential matrix [55] for relative motion between two perspective cameras ($[\mathbf{t}]_\times \mathsf{R}$). This implies that we can map the axis estimation to the 5-point algorithm for essential matrix computation [103]. Given 8 correspondences, we obtain an 8×12 matrix B as above. Let $\mathbf{V}_{i=1}^4$ be the right null singular vectors of B. The solution lies in a four dimensional subspace

$$\begin{bmatrix} \mathsf{E}(:) \\ \mathbf{s} \end{bmatrix} = \lambda_1 \mathbf{V}_1 + \lambda_2 \mathbf{V}_2 + \lambda_3 \mathbf{V}_3 + \lambda_4 \mathbf{V}_4, \tag{4.4}$$

where the λ_i's are unknown scalars. λ_4 can be set to 1 since the solution can be recovered only upto a scale factor anyhow. The 'E' part of the solution is

$$\mathsf{E}(:) = \lambda_1 \mathbf{V}_1(1:9) + \lambda_2 \mathbf{V}_2(1:9) + \lambda_3 \mathbf{V}_3(1:9) + \mathbf{V}_4(1:9), \tag{4.5}$$

where $\mathbf{V}_i(1:9)$ denote the first 9 elements of $\mathbf{V}_i$. Now the λ_i's can be computed using the solution in [103] by providing the above subspace vectors for E.

After recovering E and $\mathbf{s}$, the axis is computed as the left null singular vector of E (since $\hat{\mathbf{a}}^\top \mathsf{E} = 0$).

The sign ambiguity in the axis is resolved by letting it point away from the camera. The translation orthogonal to the axis, $\mathbf{t}_{A^\perp}$, can be obtained as $\mathbf{s} \times \hat{\mathbf{a}}$. Four solutions for R are recovered from E as in [103]. The correct solution can be recovered along with the other unknown parameters like layer thickness by analyzing the feasibility of different light-path setups as shown in Sections 4.6 Interestingly, the axis estimation is similar to the center of distortion estimation for central cameras in [54] and our 8pt algorithm can be applied.

The axis computation approach mentioned thus far is really a partial calibration of the axial camera setup mentioned earlier. While the 8pt algorithm requires generic 3D points in order to give unique solutions, it would be useful to find an alternate approach that could work with planar 3D correspondences for a variety of reasons like ease of capture, fewer images etc..

Planar Grid Calibration We now show that calibration can also be done using a single planar grid, which is useful in practice. We describe an 8pt algorithm as follows. Without loss of generality, assume the plane is aligned with the xy plane ($\mathsf{X}^i_z = 0$). Substituting in the coplanarity equation 4.2, the columns 7, 8, 9 of B matrix reduce to zero. Let B' be the reduced 8×9 matrix, whose rank is 8. Thus, there is a single right null singular vector, contrary to the above case for the original full 8×12 matrix B. This singular vector directly gives the first two columns of E, along with $\mathbf{s}$. The last column of E is recovered using Demazure constraints [29] and $\det(\mathsf{E}) = 0$ constraint, as follows.

Let $\mathsf{E} = \begin{bmatrix} e_1 & e_4 & x \\ e_2 & e_5 & y \\ e_3 & e_6 & z \end{bmatrix}$, where the e_i's are estimated as above and x, y, z are unknown. Setting $\det(E) = 0$ gives a linear equation in x, y, z which can be written as

$$xe_2e_6 - xe_3e_5 - ye_1e_6 + ye_3e_4 + ze_1e_5 - ze_2e_4 = 0 \tag{4.6}$$

using which x can be obtained in terms of y and z as

$$x = ((e_1e_6 - e_3e_4)y + (e_2e_4 - e_1e_5)z)/(e_2e_6 - e_3e_5) \tag{4.7}$$

Let

$$K = e_1^2 + e_2^2 + e_3^2 + e_4^2 + e_5^2 + e_6^2 + x^2 + y^2 + z^2 \tag{4.8}$$

The Demazure constraints arise from the form of E (= $[.]_\times$R) [29], and give the following nine equations

$$\begin{aligned}
x(2e_1^2+2e_4^2+2x^2)-xK+y(2e_1e_2+2e_4e_5+2xy)& \\
+z(2e_1e_3+2e_4e_6+2xz) &= 0 \qquad (4.9)\\
y(2e_2^2+2e_5^2+2y^2)-yK+x(2e_1e_2+2e_4e_5+2xy)& \\
+z(2e_2e_3+2e_5e_6+2yz) &= 0 \qquad (4.10)\\
z(2e_3^2+2e_6^2+2z^2)-zK+x(2e_1e_3+2e_4e_6+2xz)& \\
+y(2e_2e_3+2e_6e_6+2yz) &= 0 \qquad (4.11)\\
e_1(e_1^2+e_4^2+x^2)-e_1K+e_2(e_1e_2+e_4e_5+xy)+e_3(e_1e_3+e_4e_6+xy) &= 0 \qquad (4.12)\\
e_4(e_1^2+e_4^2+x^2)-e_4K+e_5(e_1e_2+e_4e_5+xy)+e_6(e_1e_3+e_4e_6+xy) &= 0 \qquad (4.13)\\
e_2(e_2^2+e_5^2+y^2)-e_5K+e_1(e_1e_2+e_4e_5+xy)+e_3(e_2e_3+e_5e_6+yz) &= 0 \qquad (4.14)\\
e_4(e_2^2+e_5^2+y^2)-e_5K+e_4(e_1e_2+e_4e_5+xy)+e_6(e_2e_3+e_5e_6+yz) &= 0 \qquad (4.15)\\
e_4(e_3^2+e_6^2+z^2)-e_6K+e_1(e_1e_3+e_4e_6+xz)+e_2(e_2e_3+e_5e_6+yz) &= 0 \qquad (4.16)\\
e_4(e_3^2+e_6^2+z^2)-e_6K+e_4(e_1e_3+e_4e_6+xz)+e_5(e_2e_3+e_5e_6+yz) &= 0 \qquad (4.17)
\end{aligned}$$

Note that the first three equations have cubic terms of x, y, z while the next six equations have quadratic terms. We can choose any two of these six quadratic equations. Let us choose the first two of the six quadratic equations and denote them as EQ_2 and EQ_3. Substituting x equation (4.7) we get two equations of the following form

$$\begin{aligned}
EQ_2 &: k_{11}y^2+k_{12}yz+k_{13}z^2+k_{14}=0, \qquad (4.18)\\
EQ_3 &: k_{21}y^2+k_{22}yz+k_{23}z^2+k_{24}=0 \qquad (4.19)
\end{aligned}$$

where the k_{ij} depend on the e_i's and are known coefficients. We can eliminate y^2 from the above two equations to get y in terms of z

$$y=\frac{k_{21}(k_{13}z^2+k_{14})-k_{11}(k_{23}z^2+k_{24})}{k_{11}k_{22}z-k_{12}k_{21}z} \qquad (4.20)$$

Substituting y back into EQ_3 gives a fourth degree equation in z

$$g_1 z^4 + g_2 z^2 + g_3 = 0 \tag{4.21}$$

where

$$\begin{aligned} g_1 &= k_{11}(k_{11}^2 k_{23}^2 - k_{11}k_{12}k_{22}k_{23} - 2k_{11}k_{13}k_{21}k_{23} + k_{11}k_{13}k_{22}^2 + k_{12}^2 k_{21}k_{23} - \\ &\quad k_{12}k_{13}k_{21}k_{22} + k_{13}^2 k_{21}^2) \qquad (4.22) \\ g_2 &= k_{11}(k_{11}k_{14}k_{22}^2 + 2k_{13}k_{14}k_{21}^2 + k_{12}^2 k_{21}k_{24} + 2k_{11}^2 k_{23}k_{24} - k_{11}k_{12}k_{22}k_{24} - \\ &\quad 2k_{11}k_{13}k_{21}k_{24} - 2k_{11}k_{14}k_{21}k_{23} - k_{12}k_{14}k_{21}k_{22}) \qquad (4.23) \\ g_3 &= k_{11}(k_{11}k_{24} - k_{14}k_{21})^2 \qquad (4.24) \end{aligned}$$

Note that since the above equation has only z^4 and z^2 terms, we can substitute $\gamma = z^2$ and get a quadratic equation in γ. In our experiments, we see that there are two real solutions and two imaginary solutions for z, where the real solutions differ in sign. Thus, we obtain a pair of E matrices which differ in the sign of their last column. Each pair of obtained rotation matrices also have the same property. The correct rotation matrix is chosen by checking for the determinant value of one (The determinant of incorrect rotation matrices equals -1, corresponding to a reflection).

Summary We showed that multiple flat refractive layers correspond to an *axial* system, and that in such a system all light-paths are planar in nature. We presented two approaches to compute the axis and some parameters of the transformation between object and camera coordinate systems. The major advantage in axial systems is that the resulting light-paths can be analyzed in 2D rather than 3D, and this reduces the complexity of the problems considerably. This is done by projecting all 3D quantities of a single light-path onto its POR. In the next sections, we will use this property to analyze 3 kinds of axial systems: planar refractive, planar reflective and spherical reflective systems. Before we proceed, however, we analyze the error bounds of our approach. This is done in two steps. In the first step we show simulation results that measure the noise resilience of our 8pt and 11pt algorithms, since a theoretical analysis turns out to be highly complex. We show that the 8pt algorithm gives encouraging bounded estimates of the axis and pose parameters. Given a bounded axis estimate, we then analyze the accuracy of computing the POR normal. It requires propagating an

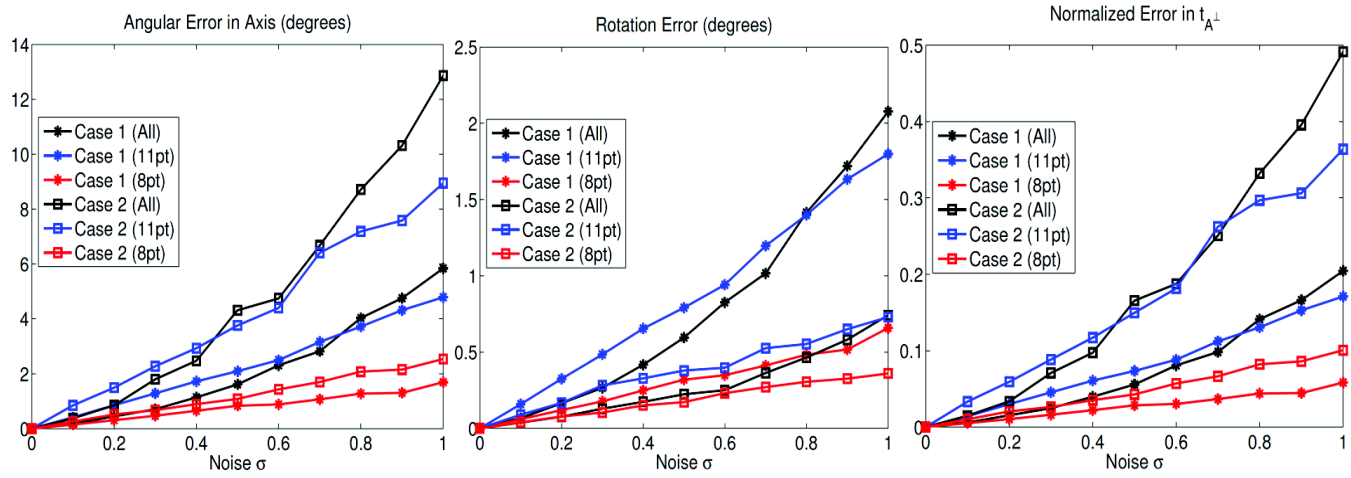


Figure 4.3: Comparison of 11pt algorithm, 8pt algorithm and using all points in a least squares fashion for estimation of axis, rotation and $\mathbf{t}_{A^\perp}$ using POR constraints.

error estimate through a cross product, and we derive the error bounds on the propagated quantity. While we will see that propagating this error through the remaining solutions to be presented in this chapter is difficult, knowledge of this error could provide useful information to produce more robust algorithms in the future.

4.2.2 Simulations

We present simulations for estimating the axis with Gaussian noise (variance σ^2 pixels) in feature points for Case 1 and Case 2, shown in Figure 4.3. We assume a camera with FOV of 45° and resolution 1000^2 pixels. The scene is set by choosing $d_0 = 300$ units, $d_1 = 450$ units, $\mu_1 = 1.5$ and $\mu_2 = 1$, where the variables d_0, d_1 represent the perpendicular distance between the camera center and the two layers of refraction, while the variables μ_1, μ_2 represent relative refractive index. We perform 100 trials for each noise setting, and plot the average error in axis, rotation and $\mathbf{t}_{A^\perp}$ in Figure 4.3. For each trial, the axis is randomly generated in a cone of half-angle 45° around the camera's optical axis. 100 3D points are randomly generated along with R and $\mathbf{t}$ so that they lie within $[d_0, 2d_0]$ after the last layer. For each trial, a RANSAC based framework is employed for both 11pt and 8pt algorithms using 200 iterations to choose the best solution using the POR error. Rotation error is defined as the minimum angle of rotation required to go from the estimated rotation to the

true rotation. Similarly, axis error is defined as the angle between the estimated and the true axis. The translation error is computed as the norm of the translation error vector and is normalized using the corresponding layer thickness.

Degenerate case Notice that when $\mathbf{l}||\mathbf{n}$, the POR constraint cannot be used. This is an inherent degenerate case in all axial systems. It also induces a computational bottleneck since light-paths that are close to the line parallel to the axis and passing through the camera center are unreliable for axis computation. As expected, the 8pt algorithm performs significantly better than the 11pt algorithm as well as using all points in a least squares fashion.

4.2.3 Error Bound: Pairwise computation

We showed earlier how the POR constraint could be used to obtain axis and partial pose parameters. Once the axis is computed, the POR normal corresponding to a particular image ray $\mathbf{l}$ can be computed as $\mathbf{l} \times \hat{\mathbf{a}}$. Note that because of errors, the computed value of the axis $\hat{\mathbf{a}}_\delta$ might be different from the ground truth $\hat{\mathbf{a}}$. Assuming no error in $\mathbf{l}$, we would like to know the error in the computation of the normal of POR Π.

To start the derivation, let us observe that both the quantities $\mathbf{l} \times \hat{\mathbf{a}}$ and $\mathbf{l} \times \hat{\mathbf{a}}_\delta$ lie in the plane that is perpendicular to $\mathbf{l}$. Let us call this plane Θ, while noting that this is not the plane of reflection or refraction. Now let us list the other vectors that lie in this plane.

$$\mathbf{l} \times (\hat{\mathbf{a}} \times \mathbf{l}) \quad \rightarrow \quad \mathbf{l} \cdot (\mathbf{l} \times (\hat{\mathbf{a}} \times \mathbf{l})) = 0 \tag{4.25}$$

$$\hat{\mathbf{a}}_\delta - (\mathbf{l} \cdot \hat{\mathbf{a}}_\delta)\mathbf{l} \quad \rightarrow \quad \left(\hat{\mathbf{a}}_\delta - (\mathbf{l} \cdot \hat{\mathbf{a}}_\delta)\mathbf{l}\right) \cdot \mathbf{l} = 0, (||\mathbf{l}|| = 1) \tag{4.26}$$

Let us also note that the second vector in the equations above is *perpendicular* to $\mathbf{l} \times \hat{\mathbf{a}}_\delta$ while the first vector is clearly perpendicular to $\mathbf{l} \times \hat{\mathbf{a}}$. Proving the first claim is trivial.

$$\left(\hat{\mathbf{a}}_\delta - (\mathbf{l} \cdot \hat{\mathbf{a}}_\delta)\mathbf{l}\right) \cdot (\mathbf{l} \times \hat{\mathbf{a}}_\delta) = \underbrace{(\hat{\mathbf{a}}_\delta) \cdot (\mathbf{l} \times \hat{\mathbf{a}}_\delta)}_{=0} - (\mathbf{l} \cdot \hat{\mathbf{a}}_\delta)\underbrace{(\mathbf{l} \cdot (\mathbf{l} \times \hat{\mathbf{a}}_\delta))}_{=0} \tag{4.27}$$

$$= 0 \tag{4.28}$$

Thus, we finally get the means to measure the angle between the two cross products.

$$\angle(\mathbf{l}\times\hat{\mathbf{a}},\mathbf{l}\times\hat{\mathbf{a}}_\delta)=\angle(\mathbf{l}\times(\mathbf{l}\times\hat{\mathbf{a}}),\hat{\mathbf{a}}_\delta-(\mathbf{l}\cdot\hat{\mathbf{a}}_\delta)\mathbf{l}) \tag{4.29}$$

A measure of this angle is given by the dot product of the two vectors.

$$\begin{aligned}
(\mathbf{l}\times\hat{\mathbf{a}})\times\mathbf{l} &= \hat{\mathbf{a}}-\mathbf{l}(\mathbf{l}\cdot\hat{\mathbf{a}}) && (4.30)\\
(\mathbf{l}\times(\hat{\mathbf{a}}\times\mathbf{l}))\cdot\left(\hat{\mathbf{a}}_\delta-(\mathbf{l}\cdot\hat{\mathbf{a}}_\delta)\mathbf{l}\right) &= (\hat{\mathbf{a}}-\mathbf{l}(\mathbf{l}\cdot\hat{\mathbf{a}}))\cdot\left(\hat{\mathbf{a}}_\delta-(\mathbf{l}\cdot\hat{\mathbf{a}}_\delta)\mathbf{l}\right) && (4.31)\\
&= \hat{\mathbf{a}}\cdot\hat{\mathbf{a}}_\delta-(\mathbf{l}\cdot\hat{\mathbf{a}}_\delta)(\mathbf{l}\cdot\hat{\mathbf{a}})\\
&\quad -(\mathbf{l}\cdot\hat{\mathbf{a}}_\delta)(\mathbf{l}\cdot\hat{\mathbf{a}})+(\mathbf{l}\cdot\hat{\mathbf{a}}_\delta)(\mathbf{l}\cdot\hat{\mathbf{a}}) && (4.32)\\
&= \hat{\mathbf{a}}\cdot\hat{\mathbf{a}}_\delta-(\mathbf{l}\cdot\hat{\mathbf{a}}_\delta)(\mathbf{l}\cdot\hat{\mathbf{a}}) && (4.33)\\
&= \cos(\delta)-(\mathbf{l}\cdot\hat{\mathbf{a}}_\delta)(\mathbf{l}\cdot\hat{\mathbf{a}}) && (4.34)\\
&\approx 1-(\mathbf{l}\cdot\hat{\mathbf{a}}_\delta)(\mathbf{l}\cdot\hat{\mathbf{a}}) && (4.35)
\end{aligned}$$

where $\approx$ denotes the fact that $\cos(\delta)$ is approximately equal to 1 for small values of δ. In the above equation, vectors have not been normalized, and so the actual value of the cosine will be given as

$$\begin{aligned}
\cos(\angle(\mathbf{l}\times\hat{\mathbf{a}},\mathbf{l}\times\hat{\mathbf{a}}_\delta)) &= \frac{\cos(\delta)-(\mathbf{l}\cdot\hat{\mathbf{a}}_\delta)(\mathbf{l}\cdot\hat{\mathbf{a}})}{\sqrt{1-(\mathbf{l}\cdot\hat{\mathbf{a}})^2}\sqrt{1-(\mathbf{l}\cdot\hat{\mathbf{a}}_\delta)^2}} && (4.36)\\
&\approx \frac{1-(\mathbf{l}\cdot\hat{\mathbf{a}}_\delta)(\mathbf{l}\cdot\hat{\mathbf{a}})}{\sqrt{1-(\mathbf{l}\cdot\hat{\mathbf{a}})^2}\sqrt{1-(\mathbf{l}\cdot\hat{\mathbf{a}}_\delta)^2}} && (4.37)
\end{aligned}$$

So far we have derived an equation to find the cosine of the angle between $\mathbf{l}\times\hat{\mathbf{a}}$ and $\mathbf{l}\times\hat{\mathbf{a}}_\delta$. The next step is to bound this angle by indirectly bounding the cosine value. To do this let us observe that cosine and sine are monotonic functions in the domain $0<\theta<\pi/2$. Thus assuming that the angle between $\mathbf{l}$ and $\hat{\mathbf{a}}$ is θ, we need to bound the value of $\mathbf{l}\times\hat{\mathbf{a}}_\delta$. Observe that this assumption can be used to re-write equation 4.36 as

$$\cos(\angle(\mathbf{l}\times\hat{\mathbf{a}},\mathbf{l}\times\hat{\mathbf{a}}_\delta))=\frac{\cos(\delta)-\cos(\theta)\cos(\theta_\delta)}{\sin(\theta)\sin(\theta_\delta)} \tag{4.38}$$

Now, observe that θ_δ can lie in the range, $\theta-\delta<\theta_\delta<\theta+\delta$. Also observe that in the extreme

cases the above equation takes the value of 1, which means that the minimum angle between the two vectors is 0. In order, to find the maximum value of the angle, we differentiate the above equation w.r.t θ_δ and equate it to 0.

$$\frac{\partial}{\partial \theta_\delta} \cos(\angle(\mathbf{l} \times \hat{\mathbf{a}}, \mathbf{l} \times \hat{\mathbf{a}}_\delta)) = 0 \tag{4.39}$$

$$\Rightarrow \frac{\cos(\theta)\sin(\theta_\delta)}{\sin(\theta)\sin(\theta_\delta)} - \frac{(\cos(\delta) - \cos(\theta)\cos(\theta_\delta))\sin(\theta)\cos(\theta_\delta)}{\sin^2(\theta)\sin^2(\theta_\delta)} = 0 \tag{4.40}$$

$$\Rightarrow \frac{\cos(\theta)}{\sin(\theta)} - \frac{(\cos(\delta)\cos(\theta_\delta) - \cos(\theta)\cos^2(\theta_\delta))}{\sin(\theta)\sin^2(\theta_\delta)} = 0 \tag{4.41}$$

$$\Rightarrow \frac{\cos(\theta)}{\sin(\theta)} - \frac{(\cos(\delta)\cos(\theta_\delta) - \cos(\theta) + \cos(\theta)\sin^2(\theta_\delta))}{\sin(\theta)\sin^2(\theta_\delta)} = 0 \tag{4.42}$$

$$\Rightarrow \frac{\cos(\theta)}{\sin(\theta)} - \frac{(\cos(\delta)\cos(\theta_\delta) - \cos(\theta))}{\sin(\theta)\sin^2(\theta_\delta)} - \frac{\cos(\theta)}{\sin(\theta)} = 0 \tag{4.43}$$

$$\Rightarrow \cos(\delta)\cos(\theta_\delta) = \cos(\theta) \tag{4.44}$$

$$\Rightarrow \cos(\theta_\delta) = \frac{\cos(\theta)}{\cos(\delta)} \tag{4.45}$$

Substituting this value back in equation 4.36 we get the maximum value of the sine of this angle as

$$\min_{\theta_\delta} \cos(\angle(\mathbf{l} \times \hat{\mathbf{a}}, \mathbf{l} \times \hat{\mathbf{a}}_\delta)) = \frac{\sqrt{\cos^2(\delta) - \cos^2(\theta)}}{\sin(\theta)} \tag{4.46}$$

$$= \frac{\sqrt{\sin^2(\theta) - \sin^2(\delta)}}{\sin(\theta)} \tag{4.47}$$

$$\max_{\theta_\delta} \sin(\angle(\mathbf{l} \times \hat{\mathbf{a}}, \mathbf{l} \times \hat{\mathbf{a}}_\delta) = \sqrt{1 - \frac{(\sin^2(\theta) - \sin^2(\delta))}{\sin^2(\theta)}} \tag{4.48}$$

$$= \frac{\sin(\delta)}{\sin(\theta)} \tag{4.49}$$

$$\approx \frac{\delta}{\sin(\theta)} \tag{4.50}$$

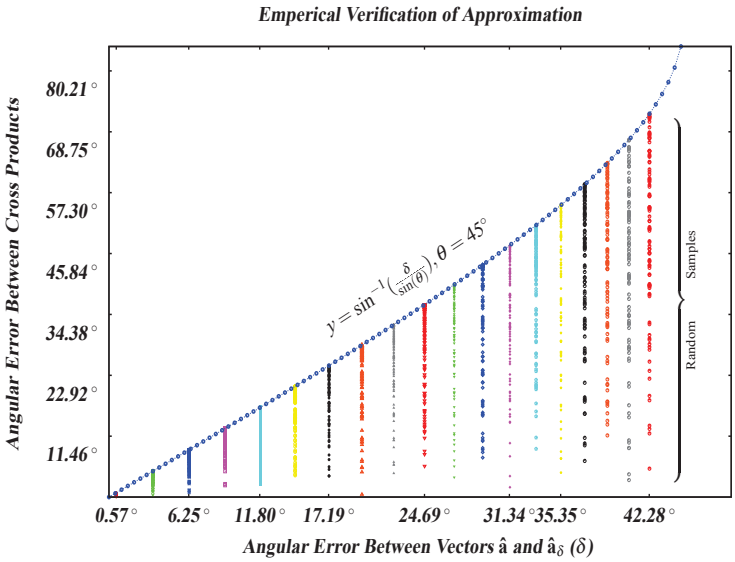


Figure 4.4: This figure presents the simulation results that confirm the bounds of our approximation theory. Two random 3D unit vectors were taken, and one of them was perturbed with several varying degrees. The resulting cross product of the two vectors was compared with the ground truth cross product, and the angle between them was measured. In the above figure, each colored dot represents the varying result plotted versus the ground truth angle between the two vectors. The curve that represents the theoretical bound is plotted to show the validity of our theory.

Finally, we get the angle bounds as

$$0 \leq \angle(\mathbf{l} \times \hat{\mathbf{a}}, \mathbf{l} \times \hat{\mathbf{a}}_\delta) \leq \sin^{-1}(\frac{\delta}{\sin(\theta)}) \tag{4.51}$$

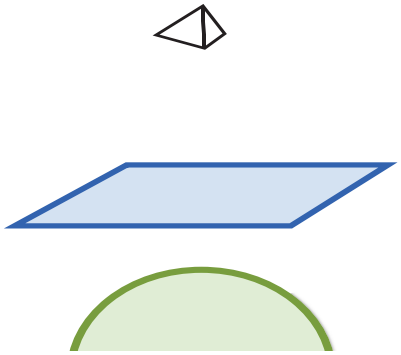
To summarize, we have bounded the angle between the vectors $\mathbf{l} \times \hat{\mathbf{a}}$ and $\mathbf{l} \times \hat{\mathbf{a}}_\delta$ where δ is the angular error in computation of the axis and consequently, the angle between these two vectors represents the angle between the ground truth normal of the POR and the estimated normal. Figure 4.4 shows simulation results that confirm our theory. Note that the above derivation is useful in another sense. When the checkerboard pose is *known* in the camera coordinate system, a closed form solution to the problem of normal estimation in the case of planar reflection or refraction can be obtained using cross products. In such a case, the error in the estimated normal is bounded in the same way as mentioned here. We know look at several individual cases of reflection/refraction to solve for the remaining light-path parameters.

4.3 General Flat Refractive Geometry

Consider the general setup for flat refractive geometry as shown in Figure 4.5, where a perspective camera observes a known calibration object via N flat refraction layers. We work in the camera coordinate system with the camera center at the origin. Let $\mathbf{n}$ denote the common surface normal to all layers and $[d_i, \mu_i]$ be the thickness and refractive index of the i^{th} medium respectively. d_0 represents the distance between the camera and the first layer. Let $[\mathtt{X}^i]_{i=1}^K$ denote K 3D points on the object which are known in the object coordinate system and let $[\mathtt{R}, \mathbf{t}]$ be the unknown rigid transformation of these points.

4.3.1 Flat Refraction Constraint

Let $[\mathbf{l}_0^i, \mathbf{l}_1^i, \ldots, \mathbf{l}_n^i]$ denote the direction vectors of each segment of the corresponding light-path. (For simplicity, we drop the superscript i for now) We assume that the internal camera calibration has been done offline and hence we know the camera ray $\mathbf{l}_0^i$ for each 3D point $\mathtt{X}^i$. The last refracted ray direction $\mathbf{l}_n$ should coincide with the line joining the transformed 3D point $\mathtt{R}\mathtt{X} + \mathbf{t}$ and the refraction point $\mathtt{p}_n$ on the last layer. Thus, the following Flat Refraction Constraint (FRC) should be satisfied.



$$\text{FRC}: \quad (\mathrm{R}\mathrm{X} + \mathbf{t} - \mathrm{p}_n) \times \mathbf{l}_n = \mathbf{0}, \tag{4.52}$$

Our goal is to estimate the unknown calibration parameters $\mathbf{n}$, $[d_k]_{k=0}^{N-1}, [\mu_k]_{k=0}^{N}$ as well as the unknown pose $[\mathrm{R}, \mathbf{t}]$ given K 2D-3D correspondences $[\mathbf{l}_0^i, \mathrm{X}^i]_{i=1}^{K}$.

From Snell's law, $\mu_i \sin(\theta_i) = \mu_{i+1} \sin(\theta_{i+1})$, where θ_i is the angle between $\mathbf{l}_i$ and $\mathbf{n}$. This can be written in vector form as

$$\mathbf{l}_{i+1} = a_{i+1}\mathbf{l}_i + b_{i+1}\mathbf{n}, \tag{4.53}$$

where $a_{i+1} = \mu_i/\mu_{i+1}$ and

$$b_{i+1} = \frac{-\mu_i \mathbf{l}_i^\top \mathbf{n} - \sqrt{\mu_i^2(\mathbf{l}_i^\top \mathbf{n})^2 - (\mu_i^2 - \mu_{i+1}^2)\mathbf{l}_i^\top \mathbf{l}_i}}{\mu_{i+1}} \tag{4.54}$$

Since Snell's law only depends on the ratio of the refractive indices, we assume $\mu_0 = 1$ without loss of generality. We first derive the FRC for a single layer and a single 3D point X. The refraction point p_1 equals $-d_0\mathbf{l}_0/(\mathbf{l}_0^\top \mathbf{n})$. Substituting in 4.52, and using equation 4.53

$$(\mathrm{R}\mathrm{X} + \mathbf{t}) \times (a_1\mathbf{l}_0 + b_1\mathbf{n}) + b_1 d_0(\mathbf{l}_0 \times \mathbf{n})/(\mathbf{l}_0^\top \mathbf{n}) = \mathbf{0}, \tag{4.55}$$

After substituting for a_1 and b_1, using equation 4.54 and removing the square root term in b_1, we get an equation with second order terms of $\mathrm{R}, \mathbf{t}$ and μ_i and sixth order terms of $\mathbf{n}$. Thus, directly solving the FRC is quite difficult. More importantly, the complexity of the FRC equation increases with each additional layer and due to the square root term in each b_i. Thus, in order to solve this problem efficiently, we need to analyze the geometry of underlying rays to derive simpler constraints. While the POR constraint (Section 4.2) allows us to estimate the normal $\mathbf{n}$, we use the framework outlined above to show how the geometry of the rays could be used to estimate 5 out of 6 pose parameters *independently* of d_i's and μ_i's (Section 4.6).

4.4 General Planar Reflective Geometry

We now derive a constraint similar to the one proposed in Section 4.3. Consider the general setup for flat reflective geometry as shown in Figure 4.1, where a perspective camera observes a known

calibration object after N reflections off parallel reflective surfaces. We work in the camera coordinate system with the camera center at the origin. Let $\mathbf{n}$ denote the common surface normal to all layers and $[d_k]_{k=0}^{N-1}$ be the perpendicular distance between each successive layer and the next. d_0 represents the distance between the camera and the first layer. Let $[\mathrm{X}]_{i=1}^{K}$ denote K 3D points on the object which are known in the object coordinate system and let $[\mathrm{R}, \mathbf{t}]$ be the rigid transformation of these points.

4.4.1 Plane Reflective Constraint

As in Section 4.3.1, we denote light-path direction vectors by $\mathbf{l}_k^i$. The last reflected ray $\mathbf{l}$ should be parallel to the line joining the transformed 3D point $\mathrm{RX}+\mathbf{t}$ and the reflection point p on the last layer. Thus, the equivalent to equation 4.52 is the Plane Reflection Constraint (PRC)

$$\mathrm{PRC}: \qquad (\mathrm{RX}+\mathbf{t}-\mathrm{p})\times\mathbf{l}=0, \tag{4.56}$$

We already outlined our approach to estimate the axis and partial translation parametsr in Section 4.2 In this case, our goal is to estimate the parameters $\mathbf{n}, [d_k]_{k=0}^{N-1}$ and the unknown pose parameter $\mathbf{t}_A$ given K 2D-3D correspondences $[\mathbf{l}_0^i, \mathrm{X}^i]_{i=1}^{K}$.

Using the laws of reflection, we can write the relationship between two consecutive segments of a light-path as

$$\mathbf{l}_{k+1} = a_{k+1}\mathbf{l}_k + b_{k+1}\mathbf{n}, \tag{4.57}$$

where $a_{k+1}=-1$ and $b_{k+1}=2(\mathbf{l}_k^\top\mathbf{n})$. Substituting this in equation 4.56, we get the following constraint for PRC of one layer (with $\mathrm{p}_1 = -d_0\mathbf{l}_0/(\mathbf{l}_0^\top\mathbf{n})$)

$$\begin{aligned}
(\mathrm{RX}+\mathbf{t}-\mathrm{p}_1)\times\mathbf{l}_1 &= \mathbf{0}, && (4.58)\\
(\mathrm{RX}+\mathbf{t}-\mathrm{p}_1)\times(2(\mathbf{l}_0^\top\mathbf{n})\mathbf{n}-\mathbf{l}_0) &= \mathbf{0}, && (4.59)\\
(\mathrm{RX}+\mathbf{t})\times(2(\mathbf{l}_0^\top\mathbf{n})\mathbf{n}-\mathbf{l}_0)+d_0/(\mathbf{l}_0^\top\mathbf{n})2(\mathbf{l}_0^\top\mathbf{n})(\mathbf{l_0}\times\mathbf{n}) &= \mathbf{0}, && (4.60)\\
(\mathrm{RX}+\mathbf{t})\times(2(\mathbf{l}_0^\top\mathbf{n})\mathbf{n}-\mathbf{l}_0)+2d_0\hat{\Pi}\mathrm{I} &= \mathbf{0} && (4.61)
\end{aligned}$$

where $\hat{\Pi}\mathrm{I}$ is the direction vector of the POR. When R, $\mathbf{n}$ and $\mathbf{t}_{A^\perp}$ are known, notice that the above equation is linear in unknown quantities d_0 and α. Also note that recursively expanding the definition

of $\mathbf{l}_{k+1}$ would give us the following

$$\begin{aligned}
\mathbf{l}_{k+1} &= -\mathbf{l}_k + 2(\mathbf{l}_k^\top \mathbf{n})\mathbf{n} && (4.62)\\
&= -(-\mathbf{l}_{k-1} + 2(\mathbf{l}_{k-1}^\top \mathbf{n})\mathbf{n}) + 2(-\mathbf{l}_{k-1}^\top \mathbf{n} + 2(\mathbf{l}_{k-1}^\top \mathbf{n}))\mathbf{n}, \qquad \because \mathbf{n}^\top \mathbf{n} = 1 && (4.63)\\
&= \mathbf{l}_{k-1} && (4.64)
\end{aligned}$$

This means that there are only 2 independent PRC equations in this case. This is understandable since reflection is a symmetric property (the incident ray is a reflection of the reflected ray) and in the absence of additional parameters like refractive index in the case of refraction, the degrees of freedom are reduced. Now, consider the case of two bounces. Since $p_2 = -d_1\mathbf{l}_1/(\mathbf{l}_1^\top \mathbf{n}) - d_0\mathbf{l_0}/(\mathbf{l}_0^\top \mathbf{n})$ and $\mathbf{l}_2 = \mathbf{l}_0$, the PRC equation reduces to

$$(\mathsf{R}\mathsf{X} + \mathbf{t} + \sum_{k=0}^{1} d_k\mathbf{l}_k/(\mathbf{l}_k^\top \mathbf{n})) \times \mathbf{l}_0 = 0 \tag{4.65}$$

Notice that the term d_0 vanishes in the above equation and so we are left with 2 unknowns d_1 and α. However, once we solve for d_1 and α, we can solve for d_0 by using the following equation

$$(\mathsf{R}\mathsf{X} + \mathbf{t} - \mathrm{p}_2) = \gamma\mathbf{l}_0 \tag{4.66}$$

Finally, for N bounces the PRC equation can be solved to obtain the perpendicular distance between every alternate pair of mirrors (1 & 2, 3 & 4,..., where mirror 1 records the first bounce of a back-projected ray from the camera). For the other pairs, we could use a generalized version of the above equation. Thus, we need $K >= N/2$ correspondences to get the required number of independent equations to solve in the linear least squares sense.

Degeneracy Note that in the case of planar reflections, there is a unique case of degeneracy. If *all* the light-paths are produced by reflection in *each* mirror of the system, then the entire system becomes *single viewpoint*. Such a system has infinitely many axes passing through it, and hence the POR constraint would not apply. Thus, for planar reflection, it is necessary that different light-paths have different numbers of bounces or intercept sets of different mirrors.

4.5 General Spherical Reflective Geometry

In Section 4.4, we looked at the case of general plane reflective geometry. In this case, the equations turn out to be linear in unknowns like the translation component $\mathbf{t}_A$ and the perpendicular distances $[d_k]_{k=0}^{N-1}$. We now consider the case of general spherical reflective geometry, when a camera looks at many axially-aligned spheres with different radii $[r_k]_{k=0}^{N-1}$ with centres with distances $[d_k]_{k=0}^{N-1}$ between consecutive spheres. d_0 represents the centre of the first sphere in the order chosen. Figure 4.2 illustrates one example of such a scenario.

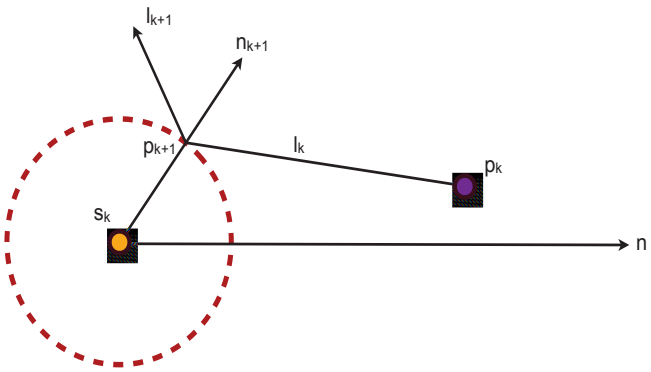
4.5.1 Spherical Reflective Constraint

As in earlier sections, the constraint we impose for any light-path is that the last segment be in the same direction as the vector from the last point p and the transformed 3D points $\mathrm{R}\mathtt{X}+\mathbf{t}$.

$$\mathtt{SRC}: \qquad (\mathrm{R}\mathtt{X}+\mathbf{t}-\mathrm{p})\times\mathbf{l}=0, \tag{4.67}$$

The main difference between the above equation and PRC, however, lies in the fact that the term $\mathbf{l}_N$ is no longer a known quantity given the axis $\mathbf{n}$. Thus, this case is similar to FRC in the event that refractive indices were unknown. As earlier, we can express the relationship between direction vectors of two consecutive segments of a light-path as (Section 4.3.1)

$$\mathbf{l}_{k+1}=a_{k+1}\mathbf{l}_k+b_{k+1}\mathbf{n}, \tag{4.68}$$


l_{k+1}
n_{k+1}
p_{k+1}
l_k
p_k
s_k
n

and the rest of the unknowns are given as

$$\begin{aligned}
\beta_1 &= -d_0\mathbf{l}_0^\top\mathbf{n} - \sqrt{d_0^2(\mathbf{l}_0^\top\mathbf{n})^2 - d_0^2 + r_1^2} && (4.74)\\
\mathsf{p}_1 &= \beta_1\mathbf{l}_0 && (4.75)\\
\mathbf{n}_1 &= (\beta_1\mathbf{l}_0 + d_0\mathbf{n})/r_1 && (4.76)\\
\mathbf{l}_1 &= 2(\beta_1 + d_0\mathbf{l}_0^\top\mathbf{n})\mathbf{n}_1/r_1 - \mathbf{l}_0 && (4.77)\\
&= 2(\beta_1 + d_0\mathbf{l}_0^\top\mathbf{n})(\beta_1\mathbf{l}_0 + d_0\mathbf{n})/r_1 - \mathbf{l}_0 && (4.78)
\end{aligned}$$

Note that after computing the axes parameters as described earlier, the unknowns in the above equation are d_0 and r_1.

Summary We have shown 3 possible scenarios in the previous section where planar axial light-paths could be solved simultaneously, and the framework in which we will solve them. Our approach essentially consists of two steps. In the first step, we use the axial nature of this configuration in order to derive most of the pose parameters relating the observed 3D points to the camera coordinate system. In the second step, we then establish the remaining parameters of the pose and light-path configuration by noting that the vector in the direction from the last bounce on the light-path towards the corresponding 3D point should be the same as the direction of the last segment of the light-path (equations FRC, PRC and SRC). Notice that while PRC is readily solvable for *any* number of bounces, solutions of FRC and SRC are restricted by the knowledge of refractive indices (FRC) and the sphere location, radius and general inter-connectedness of the light-path parameters (SRC). It is worthy noting here that the sphere is the *simplest* case of a quadric surface, and any other surface like a paraboloid or a higher order surface would only increase the amount of unknowns and render SRC even more complicated and less solvable.

4.6 Solving for Light-Paths

In the previous section, we set the framework for recovering the light-path configuration from individual observations. In this section, we solve for the remaining unknown parameters using FRC and SRC. Note that PRC has already been solved (equations 4.65,4.66). We first consider solving FRC

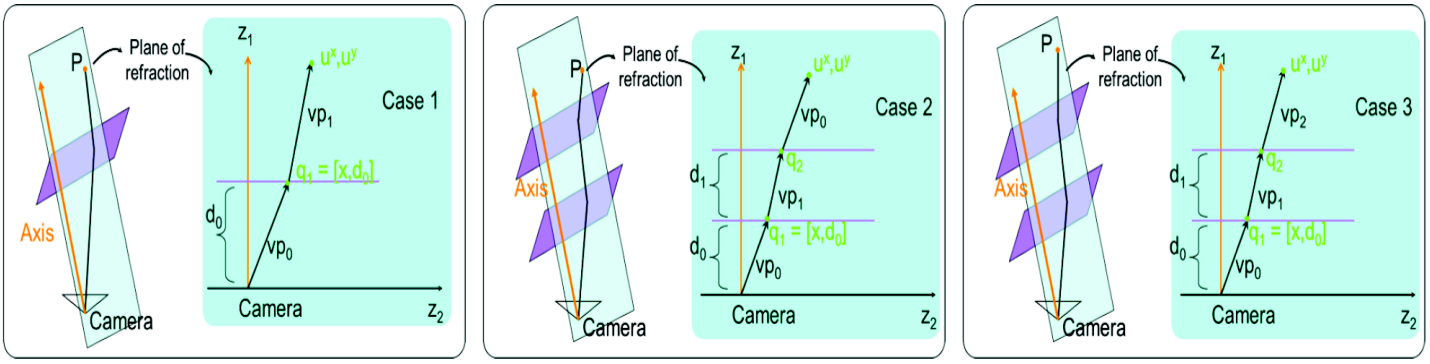


Figure 4.7: Three cases of refraction considered in this chapter. In each case we consider 4 variations: known/unknown refractive indices with known pose parameters, and known/unknown refractive indices with unknown pose parameters

in 3 different cases as shown in Figure 4.7, when the refractive indices are known and unknown. These three cases correspond to the most practical situations that arise, for example, when a camera looks through a fishtank or in underwater photography. We then look at the case of SRC for single and double bounces. Later we show that while both PRC and SRC might represent restricted scenarios of reconstruction, their applicability extends to more general cases since many smooth reflective surfaces could be approximately modeled as sets of planes or spheres. Before proceeding with the derivations, we first transform the coordinate system that we work in, for the sake of convenience.

Coordinate Transformations : We first apply the computed R and $\mathbf{t}_{A^\perp}$ to the 3D points $\mathtt{X}$. Let $\mathtt{X}_c = \mathsf{R}\mathtt{X} + \mathbf{t}_{A^\perp}$. With known axis, the analysis can be done in 2D on the plane of refraction (POR) itself as shown in Figure 4.7. Let $\mathbf{t}_A = \alpha\mathbf{A}$, where α is the unknown translation magnitude along the axis. Let $[\mathbf{z}_2, \mathbf{z}_1]$ denote an orthogonal coordinate system on the POR. We choose $\mathbf{z}_1$ along the axis. For a given camera ray $\mathbf{l}_0$, let $\mathbf{z}_2 = \mathbf{z}_1 \times (\mathbf{z}_1 \times \mathbf{l}_0)$ be the orthogonal direction. The projection of $\mathtt{X}_c$ on POR is given by $\mathbf{u} = [u^x, u^y]$, where $u^x = \mathbf{z}_2^\top \mathtt{X}_c$ and $u^y = \mathbf{z}_1^\top \mathtt{X}_c$. Similarly, the direction vector $\mathbf{l}_i$ of each ray on the light-path of $\mathbf{l}_0$ can be represented by a 2D vector $\mathbf{vp}_i$ on POR, whose components are given by $\mathbf{z}_2^\top \mathbf{v}_i$ and $\mathbf{z}_1^\top \mathbf{v}_i$. Let $c_i = \mathbf{vp}_i^\top \mathbf{z}_1$ and $\mathbf{z}_p = [0; 1]$ be a unit 2D vector.

4.6.1 Layer Thickness

We first assume known refractive indices. Section 4.2.1 showed how to compute the axis $\mathbf{A}$, rotation R and translation $\mathbf{t}_{A^\perp}$ orthogonal to the axis. When μ_i's are known, the ray directions for the entire light-path $\mathbf{v}_0(i), \ldots, \mathbf{v}_n(i)$ can be pre-computed using the estimated $\mathbf{A}$. The remaining unknowns are the layer thickness d_i's and the translation $\mathbf{t}_A$ along the axis, which can be computed *linearly* as described below.

4.6.2 Linear System for N Layers

For each correspondence, the FRC for N layer system on its plane of refraction is given by

$$\mathbf{vp}_n \times (\mathbf{u} + \alpha \mathbf{z}_p - \mathbf{q}_n) = 0. \tag{4.79}$$

This is because the last refracted ray $\mathbf{vp}_n$ should coincide with the line joining the transformed 3D point $\mathbf{u} + \alpha\mathbf{z}_p$ and the refraction point $\mathbf{q}_n$ on the last layer. $\mathbf{q}_n = \sum_{i=0}^{n-1} -d_i \mathbf{vp}_i / c_i$. Substituting, we get

$$\mathbf{vp}_n \times \begin{bmatrix} \frac{\mathbf{vp}_0}{c_0} & \cdots & \frac{\mathbf{vp}_{n-1}}{c_{n-1}} & \mathbf{z}_p \end{bmatrix} \begin{bmatrix} d_0 \\ \vdots \\ d_{n-1} \\ \alpha \end{bmatrix} = -\mathbf{vp}_n \times \mathbf{u}. \tag{4.80}$$

Thus, each correspondence gives one linear equation in d_i's and α. By stacking $K > n$ correspondences, the resulting linear system can be solved to obtain d_i's and α for n layers. After estimating α, the translation $\mathbf{t}$ is given by $\mathbf{t}_{A^\perp} + \alpha\mathbf{A}$. However, if $\mu_i = \mu_n$ for any i, $\mathbf{vp}_i \| \mathbf{vp}_n$ and d_i cannot be estimated. In addition, if $\mu_i = \mu_j$, only the combined layer thickness $d_i + d_j$ can be estimated, since the corresponding constraints in the linear system become equal. Now we analyze some special cases.

Case 1 (Single Refraction) For a single layer, we have two unknowns d_0 and α and the FRC is given by

$$\mathbf{vp}_1 \times \begin{bmatrix} \mathbf{vp}_0 / c_0 & \mathbf{z}_p \end{bmatrix} \begin{bmatrix} d_0 \\ \alpha \end{bmatrix} = -\mathbf{vp}_1 \times \mathbf{u}, \qquad \mathbf{vp}_1 = a_1 \mathbf{vp}_0 + b_1 \mathbf{n} \tag{4.81}$$

where $a_1 = 1/\mu_1$. Using $K \geq 2$ correspondences, a least squares solution can be obtained.

Case 2 (Two Refractions) $\mu_0 = \mu_2$ This is a common scenario when looking *through* a refractive medium such as a thick glass slab. Here d_0, d_1 and α are unknowns. Since $\mu_0 = \mu_2$, $\mathbf{vp}_2 \| \mathbf{vp}_0$ and the FRC reduces to

$$\mathbf{vp}_0 \times \begin{bmatrix} \mathbf{vp}_1/c_1 & \mathbf{z}_p \end{bmatrix} \begin{bmatrix} d_1 \\ \alpha \end{bmatrix} = -\mathbf{vp}_0 \times \mathbf{u} \tag{4.82}$$

Thus, we can only estimate the thickness d_1 of the medium, but not its distance d_0. This is consistent with the analysis shown in [73].

Case 3 (Two Refractions) $\mu_0 \neq \mu_2$ Now $\mathbf{v}_2$ and $\mathbf{v}_0$ are not parallel and the FRC is given by

$$\mathbf{vp}_2 \times \begin{bmatrix} \mathbf{vp}_0/c_0 & \mathbf{vp}_1/c_1 & \mathbf{z}_p \end{bmatrix} \begin{bmatrix} d_0 \\ d_1 \\ \alpha \end{bmatrix} = -\mathbf{vp}_2 \times \mathbf{u} \tag{4.83}$$

Thus, we can estimate the distance d_0 as well.

4.6.3 Unknown Refractive Indices

In this section, we describe in detail the analytical solutions to compute the layer thickness and translation along the axis when the refractive indices are unknown. We apply the same coordinate transformations as described earlier. Note that on the plane of refraction, the normal $\mathbf{n}$ of the refracting layers is given by $\mathbf{n} = [0; -1]$.

Case 1: Single Refraction We have three unknowns d_0, μ_1 and α. When μ_i's are unknown, ray directions cannot be pre-computed and FRC needs to be written in terms of camera rays as follows

$$(\alpha_1 \mathbf{vp}_0 + b_1 \mathbf{z}_1) \times (\mathbf{u} + \alpha \mathbf{z}_p + d_0 \mathbf{vp}_0/c_0) = 0 \tag{4.84}$$

Since the camera ray $\mathbf{vp}_0$ is known, we can normalize it. Let $\mathbf{vp}_0 = [v^x; v^y]$. From 4.54,

$$b_1 = \frac{v^y - \sqrt{\mu_1^2 + (v^y)^2 - 1}}{\mu_1} \tag{4.85}$$

Using a_1 and b_1, $\mathbf{vp}_1$ can be obtained. Substituting $\mathbf{vp}_1$ and $\mathbf{vp}_0$ in the FRC equation 4.81

$$(d_0 v^x - v^y u^x)\sqrt{\mu_1^2 + (v^y)^2 - 1} + v^x v^y(\alpha - d_0 - v^y) = 0 \tag{4.86}$$

Removing the square root term, we get

$$(d_0 v^x - v^y u^x)^2(\gamma + (v^y)^2 - 1) = (v^x v^y(\alpha - d_0 - u^y))^2 \tag{4.87}$$

where $\gamma = \mu_1^2$. γ can be obtained as a function of d_0 and α.

$$\gamma = \frac{(v^x v^y(\alpha - d_0 - u^y))^2}{(d_0 v^x - v^y u^x)^2} - (v^y)^2 + 1 \tag{4.88}$$

Let $[EQ_i]_{i=1}^3$ be the 3 equations for 3 correspondences. Using EQ_1, γ can be obtained as a function of d_0 and α as above. Substituting γ in EQ_2 and EQ_3 makes them cubic in d_0 and quadratic in α. We get the following form for EQ_2 and EQ_3

$$\begin{aligned} EQ_2: \quad & k_{11}\alpha^2(k_{12}d_0^2 + k_{13}d_0 + k_{14}) + k_{15}\alpha(k_{16}d_0^3 + k_{17}d_0^2 + k_{18}d_0 + k_{19}) + \\ & (k_{31}d_0^3 + k_{32}d_0^2 + k_{33}d_0 + k_{34}) = 0 \end{aligned} \tag{4.89}$$

$$\begin{aligned} EQ_3: \quad & k_{21}\alpha^2(k_{22}d_0^2 + k_{23}d_0 + k_{24}) + k_{25}\alpha(k_{26}d_0^3 + k_{27}d_0^2 + k_{28}d_0 + k_{29}) + \\ & (k_{41}d_0^3 + k_{42}d_0^2 + k_{43}d_0 + k_{44}) = 0 \end{aligned} \tag{4.90}$$

where k_{ij} depends on known quantities. α^2 can be eliminated between EQ_2 and EQ_3 by

$$EQ_4 = k_{21}k_{22}EQ_2 - k_{11}k_{12}EQ_3. \tag{4.91}$$

The resulting EQ_4 is linear in α and cubic in d_0, using which α can be obtained as a cubic function of d_0. Substituting α in EQ_3 and simplifying, results in a 6^{th} degree equation in the single unknown

d_0. Solving it results in 6 solutions. The correct solution in practice is always found by enforcing $d_0 > 0, \alpha > d_0$ and $\mu_1 > 0$. However, we have been unable to prove theoretically that this solution is unique.

Case 2: Two Refractions, $\mu_2 = \mu_0$ In this case, we have four unknowns d_0, d_1, μ_1 and α. However, d_0 cannot be estimated as shown in a similar case with known refractive index, earlier. The resulting FRC turns out to be independent of d_0. For this case, the FRC is given by

$$\mathbf{vp}_0 \times (\mathbf{u} + \alpha \mathbf{z}_p - \mathbf{q}_2) = 0 \tag{4.92}$$

since $\mathbf{vp}_2$ is parallel to $\mathbf{vp}_0$. The refraction point $\mathbf{q}_2$ is given by

$$\mathbf{q}_2 = \mathbf{q}_1 - d_1 \mathbf{vp}_1 / (\mathbf{vp}_1^\top \mathbf{n}) \tag{4.93}$$

$\mathbf{vp}_1$ is given by

$$\mathbf{vp}_1 = a_1 \mathbf{vp}_0 + b_1 \mathbf{n}, \tag{4.94}$$

where $a_1 = 1/\mu_1$. Since the camera ray $\mathbf{vp}_0$ is known, we can normalize it. Let $\mathbf{vp}_0 = [v^x; v^y]$. From 4.92,

$$b_1 = \frac{v^y - \sqrt{\mu_1^2 + (v^y)^2 - 1}}{\mu_1} \tag{4.95}$$

Using a_1 and b_1, $\mathbf{vp}_1$ and $\mathbf{q}_2$ can be obtained. Substituting $\mathbf{vp}_1$ and $\mathbf{vp}_0$ in the FRC equation 4.92

$$(d_1 v^x - \alpha v^x + v^y u^x - v^x u^y)\sqrt{\mu_1^2 + (v^y)^2 - 1} + d_1 v^x v^y = 0 \tag{4.96}$$

Removing the square root term, we get

$$(d_1 v^x - \alpha v^x + v^y u^x - v^x u^y)^2 (\gamma + (v^y)^2 - 1) = (d_1 v^x v^y)^2 \tag{4.97}$$

where $\gamma = \mu_1^2$. Once again, γ can be obtained as a function of d_1 and α.

$$\gamma = \frac{(d_1 v^x v^y)^2}{(d_1 v^x - \alpha v^x + v^y u^x - v^x u^y)^2} - (v^y)^2 + 1 \tag{4.98}$$

Similar to Case 1, let $[EQ_i]_{i=1}^{3}$ be the 3 equations for 3 correspondences. Using EQ_1, γ can be obtained as a function of d_1 and α as above. Substituting γ in EQ_2 and EQ_3 makes them cubic in d_1 and fourth degree in α. We found it difficult to solve in Matlab, due to the large number of terms. Therefore, we used an automatic generator of Gröbner basis solver [72] to obtain the final equation. It results in a 6^{th} degree equation.

Case 3: Two Refractions, $\mu_2 \neq \mu_0$ In this case, we have five unknowns d_0, d_1, μ_1, μ_2 and α. However, this case is extremely difficult to solve and we were unable to get an analytical solution. As shown, in this case the FRC will result in an equation in the above five unknowns, with fourth degree terms of each unknown. Thus, it is clear that more than two layers or multi-layer systems are quite difficult to solve for analytically and require a good initial guess for non-linear refinement, when refractive indices are unknown.

For this case, the FRC is given by

$$\mathbf{vp}_2 \times (\mathbf{u} + \alpha \mathbf{z}_p - \mathbf{q}_2) \tag{4.99}$$

since $\mathbf{vp}_2$ is *not* parallel to $\mathbf{vp}_0$. $\mathbf{vp}_2$ is given by

$$\mathbf{vp}_2 = a_2\mathbf{vp}_1 + b_2\mathbf{n} = a_2a_1\mathbf{vp}_0 + (a_2b_1 + b_2)\mathbf{n} \tag{4.100}$$

where $a_2 = \mu_1/\mu_2$ and

$$b_2 = \frac{\sqrt{\mu_1^2\left(\frac{v^y}{\mu_1} - \frac{v^y - \sqrt{\mu_1^2 + (v^y)^2 - 1}}{\mu_1}\right)^2 - \mu_1^2 + \mu_2^2} - \mu_1\left(\frac{v^y}{\mu_1} - \frac{v^y - \sqrt{\mu_1^2 + (v^y)^2 - 1}}{\mu_1}\right)}{\mu_2} \tag{4.101}$$

Using a_1, b_1, a_2, b_2, we can obtain $\mathbf{vp}_2$ and $\mathbf{q}_2$. Substituting in FRC equation 4.99, we get

$$k_1\sqrt{D_1} + k_2\sqrt{D_1D_2} + k_3\sqrt{D_2} = 0, \tag{4.102}$$

where

$$
\begin{aligned}
k_1 &= v^x v^y (d_0 - \alpha + d_1 - u^y) && (4.103)\\
k_2 &= u^x v^y - d_0 v^x && (4.104)\\
k_3 &= -d_1 v^x v^y && (4.105)\\
D_1 &= \mu_1^2 + (v^y)^2 - 1 && (4.106)\\
D_2 &= \mu_2^2 + (v^y)^2 - 1 && (4.107)
\end{aligned}
$$

Removing the square root terms, we get

$$(k_1^2 D_1 + k_3^2 D_2 - k_2^2 D_1 D_2)^2 - 4k_1^2 k_3^2 D_1 D_2 = 0 \quad (4.108)$$

which is a 4^{th} degree equation in four unknowns d_0, d_1, μ_1 and α. The above equation has up to 4^{th} degree terms of each of the unknowns. We were not able to get a polynomial equation in a single unknown using 5 correspondences. Thus, multi-layer systems require a good initial guess when μ_i's are unknown. Figure 6 shows the pose and calibration estimates for real data (Figure 5) assuming unknown μ_1 for water, which was recovered as 1.296 (relative error 2.55%).

4.6.4 Single Bounce SRC

In this section, we look at solving the SRC equation 4.67 for the case of a single sphere. Note that in this case $\mathbf{n} = [0; -1]$ and thus $\mathbf{vp}_0^\top \mathbf{n} = -c_0$, where, as noted earlier $c_i = \mathbf{vp}_0^\top \mathbf{z}_1$. To reiterate, the SRC equation for a single sphere can be written as

$$\mathbf{vp}_1 \times (\mathsf{u} + \alpha \mathbf{z}_p - \mathsf{q}_1) = 0 \quad (4.109)$$

Re-writing equations 4.74-4.78 using the 2D convention presented earlier, we get

$$\begin{aligned}
\beta_1 &= d_0c_0 - \sqrt{d_0^2c_0^2 - d_0^2 + r_1^2} = d_0c_0 - \sqrt{D} && (4.110)\\
\mathsf{q}_1 &= \beta_1 \mathbf{vp}_0 && (4.111)\\
\mathbf{n}_1 &= (\beta_1 \mathbf{vp}_0 + d_0\mathbf{n})/r_1^2 && (4.112)\\
\mathbf{vp}_1 &= 2(\beta_1 - d_0c_0)\mathbf{n}_1/r_1^2 - \mathbf{vp}_0 && (4.113)\\
&= 2(-\sqrt{D})(\beta_1 \mathbf{vp}_0 + d_0\mathbf{n}_0)/r_1^2 - \mathbf{vp}_0 \quad \text{(from equation 4.110)} && (4.114)\\
&= 2(-\sqrt{D}\beta_1 \mathbf{vp}_0 - d_0\sqrt{D}\mathbf{n})/r_1^2 - \mathbf{vp}_0 && (4.115)\\
&= 2(-\sqrt{D}\beta_1 - 1/2r_1^2)/r_1^2 \begin{bmatrix} v^x \\ v^y \end{bmatrix} - 2(\sqrt{D}d_0)/r_1^2 \begin{bmatrix} 0 \\ -1 \end{bmatrix} && (4.116)
\end{aligned}$$

after expanding for $\mathbf{vp}_0$ and $\mathbf{n}$. Expanding the second term, we get

$$\mathsf{u} + \alpha \mathbf{z}_p - \mathsf{q}_1 = \begin{bmatrix} u^x - \beta_1 v^x \\ u^y + \alpha - \beta_1 v^y \end{bmatrix} \tag{4.117}$$

Finally, substituting in the SRC equation we get

$$(-\beta_1\sqrt{D} - 1/2r_1^2)(v^x u^y + v^x\alpha - v^y u^x) - \sqrt{D}d_0(u^x - \beta_1 v^y) = 0 \tag{4.118}$$

The above equation is linear in α and so we obtain α in terms of the other unknowns d_0 and r_1 as (note that β_1 is expressed in terms of d_0 and r_1)

$$\alpha = \frac{(\beta_1\sqrt{D} + 1/2r_1^2)(v^y u^x - v^x u^y) - (\sqrt{D}d_0)(u^x - \beta_1 v^x)}{v^x(-\beta_1\sqrt{D} - 1/2r_1^2)} \tag{4.119}$$

Using another correspondence, we get another equation similar to equation 4.118 but with different coefficients. Unfortunately, substituting the value for α in the new equation results in binomial terms of higher degree in the unkowns d_0, r_1. We found it difficult to solve it in both Matlab and using Gröbner Bases. We also tried other representations (using polar coordinaes, for example) but were unable to deduce a solution for this case. Note that a higher number of spheres or shapes with more complexity will contain more unknowns, and hence will be more complicated than the equations

presented here.

4.6.5 Single SRC With Known Pose

Notice that in equation 4.118 there is only one term α that is related to the relative position between camera and 3D correspondences. Also, note that it is easy to isolate terms with r since mostly they are only contained in the term D. We used 2 correspondences to solve for r in terms of d, which results in four 4th degree equations involving only the sphere center. Thus we get 16 possible sphere centres, and for each solution of d, 4 values for r for a total of 64 solutions. Disambiguating could be done by observing some general rules like $d > 0, r > 0$, $d > r$ and using a 3rd or more correspondences.

4.6.6 Solving Planar Reflection With Known Pose

Similar to the spherical case, we can also solve the planar reflection scenario in the presence of known relative poses between camera and 3D correspondences. In such a case, the POR constraint has only 1 unknown in the plane normal, and hence can be solved using 2 correspondences. As for the PRC, there is a reduction of only one variable α, and so the solution is more or less the same as described earlier.

Summary We have shown that, for planar light-paths, it is possible to exploit the axial property to compute the relative pose of 3D correspondences and the optical system and hence reconstruct the light-paths in various scenarios. We divide the problem into two parts. In the first part, we find the axis direction and part of the pose parameters using the POR constraints. In the second part, for individual cases we use the refraction or reflection constraints to compute the remaining pose and light-path parameters. In certain cases, we also showed that reconstruction was not possible. Finally, we showed that when the relative pose is known, we can reconstruct single planar and spherical reflection using 2 correspondences. Table 4.1 shows the various cases handled in this chapter.

Flat Refraction Problems

	Known Pose	Unknown Pose
Case 1	$\surd(2)/\surd(2)$	$\surd(2)/\surd(3)$
Case 2	$\surd(2)/\surd(2)$	$\surd(2)/\surd(3)$
Case 3	$\surd(3)/\surd$	$\surd(3)/\times$
Case N	$\surd(N)/\times$	$\surd(N)/\times$

Plane and Spherical Reflection Problems

	Known Pose	Unknown Pose
PRC 1 layer	$\surd(2)$	$\surd(2)$
PRC 2 layers	$\surd(2)$	$\surd(2)$
PRC N layers	$\surd(N)$	$\surd(N)$
SRC 1 layer	$\surd(2)$	$\times$

Table 4.1: Table showing tractable flat refraction problems and plane and spherical reflective problems. Each box in flat refraction indicates whether a particular scenario is solvable with / without knowledge of refractive indices. The numbers in the brackets indicate the number of correspondences involved in each minimal solution

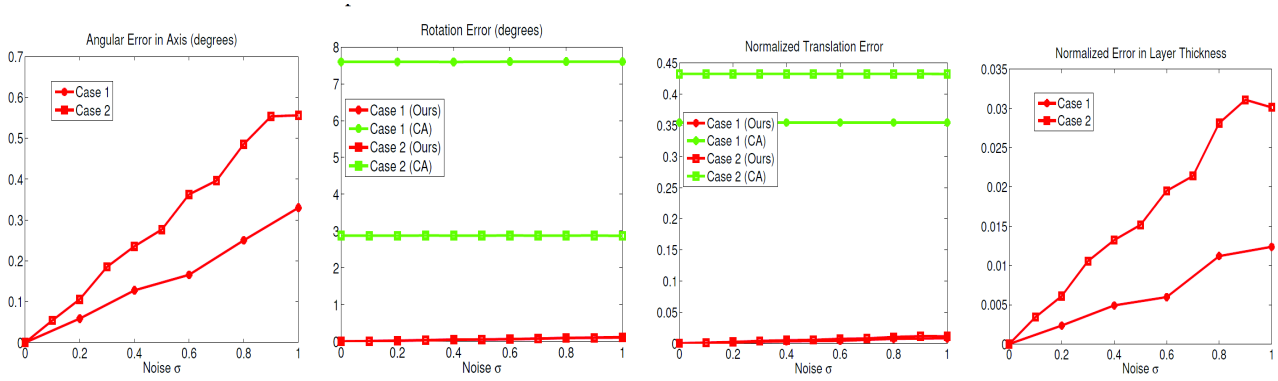


Figure 4.8: Error in axis, rotation, translation and layer thickness using a planar calibration grid for different noise values, averaged over 100 trials. Rotation and translation errors using a central approximation (CA) are also shown.

4.7 Results

In this section, we present results for various simulations that confirm our theories, as well as real results and comparisons for flat refraction.

4.7.1 Simulations

Now we present simulations for the complete calibration and pose estimation process, using the earlier settings. The 8pt algorithm is used since it works better. In the RANSAC framework, after

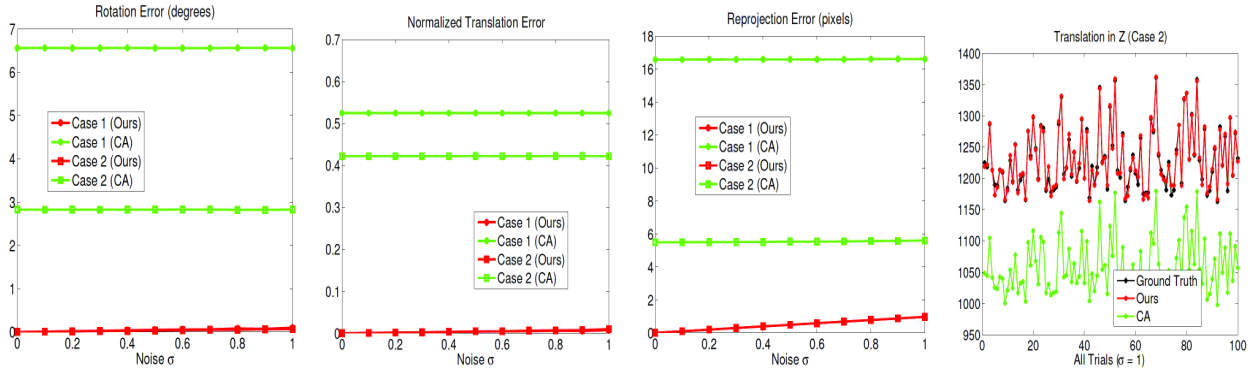


Figure 4.9: Rotation, translation and reprojection error using our algorithm versus using a central approximation (CA) for Case 1 and Case 2. The right most plot shows the estimated $\mathbf{t}_z$ for Case 2 over all 100 trials for $\sigma = 1$ pixel. CA estimates the object to be closer to the camera than in reality.

estimating the axis, the best set of 8 points are used to computed α and d_i's in a least square fashion as described earlier. Since there are 4 solutions for R from E matrix, we get 4 solutions for α and d_i's. The correct solution is found by enforcing $\alpha > \sum d_i$ and $d_i > 0 \forall i$. We also compute the pose obtained using a central (perspective) approximation from the given 2D-3D correspondences (referred by CA). Figures 4.8 shows error plots for pose and calibration parameters and the final reprojection error for different noise levels, averaged over 100 trials. These plots show that correct calibration and pose parameters can be obtained using our algorithm. Notice the large translation and reprojection error, and smaller estimated $\mathbf{t}_z$ when using a central approximation (CA) in Figure 4.8. This is expected because when looking through a denser refractive medium, objects appear *closer* to the camera. Note that the error due to noise in CA is insignificant compared to the error due to incorrect modeling.

4.7.2 Real Results

We show real results using a water tank of dimensions $508 \times 260 \times 300mm^3$. We use a Canon Rebel XT camera having resolution of 3456×2304 pixels with a $18 - 55$ mm zoom lens. The camera was internally calibrated offline. Figure 4.10 shows a photo of a scene consisting of three checkerboards, captured by looking through the water tank (facing $260mm$ side of tank). In order to obtain ground truth, we took another photo in air, using which the poses of the of checkerboards were computed. The resulting 3D points in the coordinate system of the left checkerboard are shown in

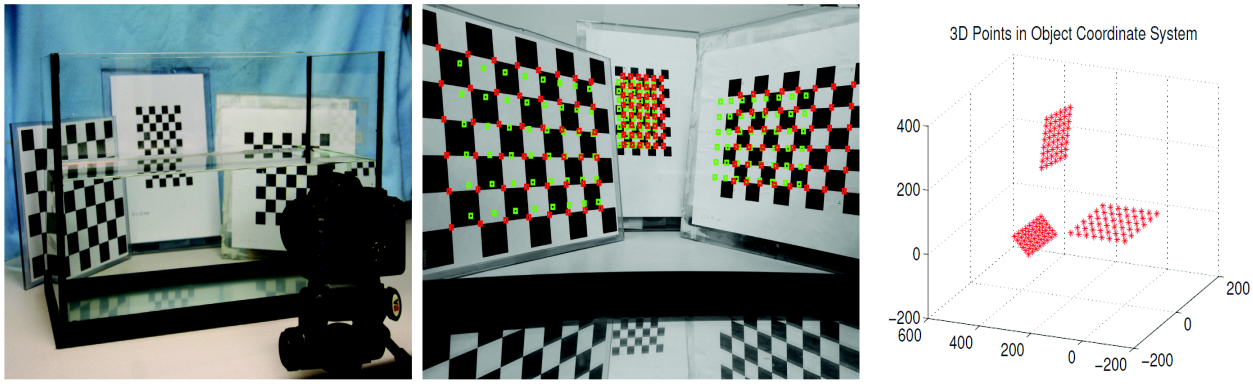


Figure 4.10: (Left) Setup. (Middle) Photo captured by looking through a water tank. Projected 3D points are overlayed by applying pose estimated using CA (green) and our algorithm (red).(Right) Reconstructed 3D points.

	N	$\theta_x, \theta_y, \theta_z$ (deg)	t_x, t_y, t_z (mm)	d_1(mm)
GT		131.38, 1.22, 84.07	-237.58, -128.85, 455.80	260
CA	144	130.24, 1.42, 83.84	-217.71, -120.73, 372.14	-
Ours[1,2]	144	131.38, 1.26, 84.12	-237.11, -128.16, 453.12	255.69
Ours[1,3]	48	131.40, 1.36, 84.03	-239.76, -129.26, 456.34	272.81
Ours[4]	144	131.37, 1.26, 84.12	-236.46, -127.86, 449.70	262.39

Table 4.2: Estimates of pose and water-tank thickness d_1 for real data shown in Figure 4.10 using central approximation (CA) and our algorithm. GT denotes ground truth and N denotes number of 2D-3D correspondences. Conditions: [1] Assuming $\mu_1 = 1.33$. [2] All Planes. [3] Left Plane Only. [4] All Planes, unknown μ_1.

Figure 4.10(Right). We detect corners in the captured photo and run our algorithm (Case 2) to estimate the calibration and pose parameters. The estimated thickness of the tank using our algorithm was 255.69mm, resulting in a relative error of 1.66%. Notice the large error in $\mathbf{t}_z$ in the central approximation (Table 4.2), also evident from projected points in Figure 4.10. Interestingly, the central approximation can recover the rotation well enough.

4.8 Discussion

We have analyzed the geometry of a perspective camera imaging through multiple flat refractive layers. We developed a theory for calibration and extended it to the case of planar and spherical reflections, which can be directly used in applications such as 3D reconstruction [9]. We presented a comprehensive analysis under unknown layer distances and orientation, and known/unknown refractive indices. Since calibration can be done using a single planar grid, the proposed algorithms are useful in practical scenarios such as underwater imaging. Multiple planar grids can be used to increase the calibration accuracy similar to calibration of perspective cameras. Our proposed 8-point algorithm for axis computation can be used for other axial setups such as catadioptric cameras, as well as to compute the distortion center for fish-eye cameras. Developing a minimal solution for calibrating flat refractive geometry remains an interesting future work.

5

Conclusion and Future Work

In this chapter we conclude this thesis on shape estimation for specular surfaces. We first summarize the main contributions of our work and then discuss possible future directions.

5.1 Summary and Discussion

Shape estimation for specular surfaces can be considered one of the last frontiers of the shape estimation problem, since for most other objects reasonably accurate algorithms exist under appropriate conditions. Because of this fact, in this thesis, we have looked at different approaches to the specular shape estimation problem. While centering around the light-path triangulation framework, we try to use different conditions and inputs to derive the following algorithms for shape estimation

- In Chapter 2, we derive the multiple view geometry of flat refraction. We consider the case where the camera and scene are in different medium, separated by a flat transparent refrac-

tive surface. We analyze the multiple view geometry setup arising from such a scenario, and present useful representations for the camera projection matrix, the fundamental matrix and the homography matrix. We also show some conditions under which the normal to this surface might be easily estimated. The underlying philosophy of this approach was to gain a geometric understanding of image formation under flat refraction, and our work shows that under such circumstances it is beneficial to group the light-paths corresponding to the image of a 3D line into one entity in order to analyze the situation.

- In Chapter 3, we explore the idea of adding photometric information to the light-path triangulation framework. While the resulting geometric analysis framework remains similar to that of [73], we show that addition of this information results in reducing the minimal requirements of reconstruction. We divide our analysis into single bounce and double bounce cases, and show encouraging results and comparisons on the single bounce case. We also theoretically analyze the conditions under which double bounce reconstruction might be possible. However at present practical problems limit our ability to produce accurate results.

- In Chapter 4, we analyze a different grouping of light-paths, when the underlying system of rays captured by the camera is axial in nature. We show that under many circumstances like refraction across parallel flat layers, such a system produces planar light-paths. We analyze the properties arising from this planarity and demonstrate the possibility of going beyond the limitations propounded in [73] in some cases. While encouraging results are presented on real scenarios like viewing across parallel refractive layers, we also derive unique insights into some specular reflective setups. Comparisons with a central approximation approach show the benefits of our methods.

To summarize, there is a lot of potential in light-path analysis for shape estimation of specular surfaces. While we presented 3 approaches with encouraging results, we believe there is a lot of scope for future work in this area.

5.2 Future Work

As discussed in each of the chapters, there is large room for improvements. Based on this, we list a few potential future directions for shape estimation and discuss some potential applications.

- One of the potential drawbacks of the phenomenon of specular refraction / reflection is that given all the parameters like knowledge of 3D points, specular surface geometry and camera poses and intrinsics, it is difficult to analytically derive the projection of a selected 3D point onto the image. Primarily, this problem occurs due to the piece-wise linear nature of the light-path associated with that 3D point. While a few approaches have been presented in the past [3], applications like shape estimation or ray tracing would benefit from more research in this area. While analytical formulae generally tend to be of high order, methods like Gröbner basis could be used for arriving at hitherto unknown solutions. Of course, presence of such analytical solutions would also help in designing new optimization based reconstruction approaches for specular surfaces.

- While a lot of research has been devoted to the geometry related to specular surfaces, little has been done in other areas like appearance based analysis of transparency [75]. Especially in problems like robot navigation, such appearance based approaches are essential to isolate transparent objects for further geometric analysis if required. In fact, exploration of the relationship between appearance and geometry of a specular surface has found early interest [70], and might be of immense use for shape estimation problems.

- One of the major bottlenecks to multiview reconstruction of specular surfaces is the correspondence problem. While it is impossible to estimate the correspondence between the location of the image of a 3D point on the surface of a specular object in two camera views, an analysis of simple circumstances (like points on a sphere or quadric) might provide better information for both reconstruction and specular flow problems. The fact that many everyday specular objects' shape can be describe in terms of such simple primitives highlights the potential applications of such a theory.

In summary, in this thesis we have presented multiple approaches to shape estimation of specular surfaces. In the future, we hope that our work will find useful applications in the fields of reconstruction, ray tracing and general image understanding of specular surfaces.

Bibliography

[1] Yair Adato, Yuri Vasilyev, Ohad Beh-Shahar and Todd Zickler. Towards a theory of shape from specular flow. *IEEE International Conference on Computer Vision*, pp. 1-8, 2007. 7, 8

[2] Sameer Agarwal, Satya Mallick, David Kriegman and Serge belongie. On refractive optical flow. *European Conference on Computer Vision*, pp. 279-290, 2004.

[3] Amit Agrawal, Srikumar Ramalingam, Yuichi Taguchi and Visesh Chari. A theory of flat refractive geometry. *IEEE Conference on Computer Vision and Pattern Recognition*, 2012. 12, 114

[4] D.G. Aliaga and Y. Xu. An adaptive correspondence algorithm for modeling scenes with strong interreflections. *IEEE Transactions on Visualization and Computer Graphics*, 15:465–480, 2009. 62

[5] D.G. Aliaga and Y. Xu. A self-calibrating method for photogeometric acquisition of 3D objects. *IEEE Pattern Analysis and Machine Intelligence*, 32:747–754, 2009. 61

[6] B. Atcheson, I. Ihrke, D. Bradley, W. Heidrich, W. Magnor and H.-P. Seidel. Imaging and 3D Tomographic reconstruction of Time-varying inhomogeneous refractive index fields. *Tech Report TR-2007-06*, University of British Columbia, Jan 2007.

[7] G. A. Atkinson and E. R. Hancock. Recovery of surface orientation from diffuse polarization. *IEEE Transactions on Image Processing*, pp. 1653-1664, 2006. 11, 12

[8] Simon Baker and Shree K. Nayar. A Theory of single-viewpoint catadioptric image formation. *International Journal of Computer Vision*, pp. 175-196, 1999. 6, 7

[9] Moshe Ben-Ezra and Shree K. Nayar. What does motion reveal about transparency? In *IEEE International Conference on Computer Vision*, page 1025, Washington, DC, USA, 2003. IEEE Computer Society. ISBN 0-7695-1950-4. 22, 23, 41, 111

[10] J.-A. Beraldin. Integration of laser scanning and close-range photogrammetry - the last decade and beyond. *XXth ISPRS congress*, pp. 972-983, 2004.

[11] Dinkar Bhat and Shree Nayar. Stereo in the presence of specular reflection. *IEEE International Conference on Computer Vision*, pp. 1086-1092, 1995.

[12] R. Bhotika, David J. Fleet and Kyros Kutulakos. A probabilistic theory of occupancy and emptiness. *European Conference on Computer Vision*, pp. 112-132, 2002.

[13] Andrew Blake and G. Brelstaff. Geometry from specularity. *IEEE International Conference on Computer Vision*, pp. 297-302, 1988. 7, 8

[14] Andrew Blake. Specular stereo. *International Joint Conference on Artificial Intelligence*, pp. 973-976, 1985. 7, 8

[15] Thomas Bonfort. Reconstruction de surfaces réfléchissantes à partir d'images. *PhD Thesis, Institut National Polytechnique de Grenoble*, Feb 2006. 8, 9

[16] Tomas Bonfort and Peter Sturm. Voxel carving for specular surfaces. *IEEE International Conference on Computer Vision*, pp. 591-596, 2003. 8

[17] Tomas Bonfort, Peter Sturm and Pau Gargallo. General specular surface triangulation. *Asian Conference on Computer Vision*, pp. 872-881, January 2006. 8, 12

[18] M. Born and E. Wolf. Princples of Optics. *Permagon Press*, 1965.

[19] G. Brelstaff and Andrew Blake. Detecting specular reflections using lambertian constraints. *IEEE International Conference on Computer Vision*, 1988.

[20] V. Caglioti, P. Tadde, G. Boracchi, S. Gasparini and A. Gius. Single-image calibration of off-axis catadioptric cameras using lines. *OMNIVIS*, 2007.

[21] J. Chahl and M. Srinivasan. Reflective surfaces for panoramic imaging. *Applied Optics*, pp. 8275-8285, 1997.

[22] Y.-J. Chang and T. Chen. Multi-view 3D reconstruction for scenes under the refractive plane with known vertical direction. *IEEE International Conference on Computer Vision*, 2011. 79

[23] Visesh Chari and Peter Sturm Multiple-view geometry of the refractive plane. *British Machine Vision Conference*, 2009. 79

[24] T. Chen, H. P. A. Lensch, C. Fuchs and H.-P. Seidel. Polarization and phase-shifting for 3D scanning of translucent objects. *IEEE Conference on Computer Vision and Patternn Recognition*, pp. 1-8, 2007.

[25] T. Chen, M. Goesele, and H.-P. Seidel. Mesostructure from specularity. *IEEE Conference on Computer Vision and Pattern Recognition*, 2:1825–1832, 2006. 37

[26] Z. Chen, K.-Y. K. Wong, Y. Matsushita, X. Zhu and M. Liu. Self-calibrating depth from refraction. *IEEE International Conference on Computer Vision*, 2011. iv, 10, 79

[27] Antonio Criminisi, Sing Bing Kang, Rahul Swaminathan, Richard Szeliski and P. Anandan. Extracting layers and analyzing their specular properties using epipolar plane-image analysis. *Computer Vision and Image Understanding*, 97(1), January 2005.

[28] Brian Curless and Marc Levoy. Better optical triangulation through space time analysis. *IEEE International Conference on Computer Vision*, pp. 987-994, 1995.

[29] M. Demazure. Sur deux problemes de reconstruction. Technical Report 882, INRIA, 1988. 83, 84

[30] Yuanyuan Ding and Jingyi Yu. Recovering shape characteristics of near-flat specular surfaces. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2008.

[31] Yuanyuan Ding, Jingyi Yu and Peter Sturm. Recovering specular surfaces using curved line images. *IEEE Conference on Computer Vision and Pattern Recognition*, June 2009.

[32] A. Efros, V. Isler, J. Shi, and M. Visontai. Seeing through water. *NIPS*, 2004. 41

[33] R. Feris, R. Raskar, Kar-Han Tan and M. Turk. Specular reflection reduction with multi-flash imaging. *Computer Graphics and Image Processing*, pp. 316-321, 2004.

[34] Phillipp Flach and Hans-Gerd Maas. Vision-based techniques for refraction analysis in applications of terrestial geodesy. *International Archives of Photogrammetry and Remote Sensing*, pp. 195-201, 2000.

[35] Y. Francken, T. Cuypers, T. Mertens, J. Gielis and P. Bekaert. High quality mesostructure acquisition using specularities. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-7, June 2008. 13

[36] Fresnel equations. *http://en.wikipedia.org/wiki/Fresnel_equations*. v, 11, 40, 43

[37] Chunyu Gao and N. Ahuja. A refractive camera for acquiring stereo and super-resolution images. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2316-2323, 2006.

[38] Simone Gasparini, Peter Sturm and João Barreto. Plane-based calibration of central catadioptric cameras. *IEEE International Conference on Computer Vision*, pp. 1195-1202, 2009. 8

[39] C. Geyer and K. Daniilidis. Catadioptric camera calibration. *IEEE International Conference on Computer Vision*, pp. 398-404, 1999. 8

[40] C. Geyer and K. Daniilidis. Catadioptric projective geometry. *International Journal of Computer Vision*, pp. 223-243, 2001.

[41] C. Geyer and K. Daniilidis. Paracatadioptric camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:687-695, 2002. 8

[42] C. Geyer and K. Daniilidis. Properties of the catadioptric fundamental matrix. *European Conference on Computer Vision*, pp. 140-154, 2002.

[43] C. Geyer and K. Daniilidis. Mirrors in motion: Epipolar geometry and motion estimation. *IEEE International Conference on Computer Vision*, 2003.

[44] Georg Glaeser and Hans-Peter Schröcker. Reflections on refractions. *Journal for Geometry and Graphics*, 2000. 23, 25

[45] G. Glaeser and H.-P. Schröcker. Reflections on refractions. *Journal of Geometry and Graphics*, 4(1):1-18, 2000.

[46] A. S. Glassner. An Introduction to Ray Tracing. *Morgan Kaufmann*, 1989.

[47] J. Gluckman and Shree K. Nayar. Catadioptric stereo using planar mirrors. *International Journal on Computer Vision*, 44(1):65-79, August 2001.

[48] M. Goesele, H. P. A. Lensch, J. Lang, C. Fuchs and H.-P. Seidel. Disco - acquisition of translucent objects. *ACM SIGGRAPH*, pp. 835-844, 2004.

[49] P. Golland and R. Szeliski Stereo matching with transparency and matting. *International Journal of Computer Vision*, pp. 45-61, 1999.

[50] Nuno Gracias and José Santos-Victor. Underwater video mosaics as visual navigation maps. *Computer Vision and Image Understanding*, 79(1):66–91, 2000. ISSN 1077-3142. 22

[51] K. Graves, R. Nagarajah and P. R. Stoddart. Analysis of structured highlight stereo imaging for shape measurement of specular objects. *Optical Engineering*, August 2007.

[52] Jinwei Gu, Shree K. Nayar, Eitan Grinspun, Peter N. Belhumeur and R. Ramamoorthi, Compressive Structured Light for Recovering Inhomogeneous Participating Media. *European Conference on Computer Vision*, October 2008. v, 11, 16

[53] M. A. Halstead, B. A. Barsky, S. A. Klein, B. R. Mandell. Reconstructing curved surfaces from specular reflection patterns using spline surface fiting of normals. *ACM SIGGRAPH*, pp. 335-342, 1996.

[54] R. I. Hartley and S. B. Kang. Parameter-free radial distortion correction with center of distortion estimation. *IEEE Pattern Analysis and Machine Intelligence*. 29(8):1309-1321, August 2007. 83

[55] R. I. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision. *Cambridge University Press*, second edition, 2004. 8, 82

[56] S. W. Hasinoff and K. N. Kutulakos. Confocal stereo. *European Confeence on Computer Vision*, pp. 620-634, 2006.

[57] S. W. Hasinoff. Three-Dimensional reconstruction of fire from images. *MSc thesis, University of Toronto, Department of Computer Science*, 2002.

[58] S. Hata, Y. Saito, S. Kumamura and K. Kaida. Shape extraction of transparent object using genetic algorithm. *International Conference on Pattern Recognition*, pp. 684-688, 1996.

[59] G. Healey and T. O. Binford. Local shape from specularity. *Computer Vision, Graphics and Image Processing*, pp. 62-86, April 1988.

[60] E. Hecht Optics. *Addison Wesley*, 2002. 42, 43, 47

[61] Aaron Hertzmann and Steven M. Seitz. Shape and materials by example: a photometric stereo approach. *IEEE Conference on Computer Vision and Pattern Recognition*, 2003. 12

[62] J. Höhle. Reconstruction of the underwater object. *Photogrammetric Engineering*, 984-954, 1971. 75

[63] R. Horn and C. Johnson. *Topics in matrix analysis*. Cambridge University Press, 1991. 25, 27

[64] M.B. Hullin, M. Fuchs, I. Ihrke, H.-P. Seidel, and H.P.A. Lensch. Fluorescent immersion range scanning. *SIGGRAPH*, 2008. 14, 37, 42

[65] C.P. Huynh, A. Robles-Kelly, and E.R. Hancock. Shape and refractive index recovery from single-view polarisation images. *IEEE Conference on Computer Vision and Pattern Recognition*, 1229–1236, 2010. 11, 42

[66] Ivo Ihrke, B. Goldluecke and M. Magnor. Reconstructing the geometry of flowing water. *IEEE International Conference on Computer Vision*, pp. 1055-1060, 2005.

[67] Iho Ihrke, Kyros Kutulakos, Hendrik P. A. Lensch, Marcus Magnor and Wolfgang Heidrich. Transparent and reflective scene reconstruction. *EUROGRAPHICS STAR - State of The Art Report*, May 2008. iv, 9, 10

[68] Katsushi Ikeuchi. Determining surface orientations of specular surfaces by using photometric stereo method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 661-670, 1981. 7, 12, 17

[69] Jungho Kim, Kuk-Jin Yoon, Jun-Sik Kim and Inso Kweon. Visual SLAM by Single-Camera Catadioptric Stereo. *International Joint Conference SICE-ICASE*, pp. 2005-2009, October 2006. iv, 6, 7, 20

[70] J. J. Koenderink and A. J. van Doorn. Photometric invariants related to solid shape. *Optica Acta*, 981-996. 7, 114

[71] R. Kotowski. Phototriangulation in multi-media photogrammetry. *International Archives of Photogrammetry and Remote Sensing*, XXVII, 1988.

[72] Zuzana Kúkelová, M. Bujnak and T. Pajdla. Automatic Generator of Minimal Problem Solvers. *European Conference on Computer Vision*, 2008. 104

[73] Kyros Kutulakos and Eron Steger. A theory of refractive and specular 3d shape by light-path triangulation. In *IEEE International Conference on Computer Vision*, pages 1448–1455, Washington, DC, USA, 2005. IEEE Computer Society. ISBN 0-7695-2334-X-02. doi: http://dx.doi.org/10.1109/ICCV.2005.26. iv, v, vii, 12, 13, 14, 15, 17, 18, 22, 23, 37, 38, 39, 40, 41, 43, 47, 48, 49, 69, 70, 79, 101, 113

[74] Jean-Marc Lavest, Gérard Rives, and Jean-Thierry Lapresté. Underwater camera calibration. In *European Conference on Computer Vision*, pages 654–668, London, UK, 2000. Springer-Verlag. ISBN 3-540-67686-4. 22, 23

[75] Anat Levin, Asat Zomet and Yair Weiss. Separating reflections from a single image using local features. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 306-313, 2004. 114

[76] Maxime Lhuillier. Automatic scene structure and camera motion using a catadioptric system. *Computer Vision and Image Understanding*, 109(2), pp. 186-203, February 2008. iv, 7

[77] H. Li, R. I. Hartley, and J. Kim. A linear approach to motion estimation using generalized camera models. *IEEE Conference on Computer Vision and Pattern Recognition*, 2008. 79

[78] R. Li, H, Li, W. Zou, R. G. Smith and T. A. Curran. Quantitative photogrammetric analysis of digital underwater video imagery. *IEEE Journal of Oceanic Engineering*, 2:364-375, 1997.

[79] Y. Li, S. Lin, H. Lu, S. B. Kang, H.-Y. Shum. Multibaseline stereo in the presence of specular reflections. *IEEE International Conference on Pattern Recognition*, pp. 573-576, 2002.

[80] A. Lintu, L. Hoffman, M. Magnor, H. P. A. Lensch and H.-P. Seidel. 3D reconstruction of reflection nebulae from a single image. *Vision, Modelling and Visualization*, pp. 109-116,2007.

[81] S. Lin, Y. Lee, S. B. Kang, X. Tong and H.-Y Shum. Diffuse-specular separation and depth recovery from image sequences. *European Conference on Computer Vision*, 2002.

[82] Miaomiao Liu, Kwan-Yee Kenneth Wong, Zhenwen Dai and Zhihu Chen Pose estimation from reflections for specular surface recovery. *IEEE International Conference on Computer Vision*, 2011. 8, 12

[83] H.-G. Maas. New developments in multimedia photogrammetry. *Optical 3D Measurement Techniques III*, 1995. 6, 9, 22, 79

[84] S. P. Mallick, T. Zickler, D. Kriegman and P. Belhumeur. Beyond lambert: Reconstructing specular surfaces using color. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 619-626, 2005. 12

[85] W. Matusik, H. Pfister, R. Ziegler, A. Ngan and L. McMillan. Acquisition and rendering of transparent and refractive objects. *Eurographics Symposium on Rendering*, pp. 267-278, 2002. iv, 10, 11, 14, 20

[86] B. Micusik and Tomáš Pajdla. Autocalibration and 3D reconstruction with non-central catadioptric cameras. *IEEE Conference on Computer Vision and Pattern Recognition*, 2004.

[87] Tomoo Mitsunaga and Shree Nayar. Radiometric Self Calibration. *IEEE Conference on Computer Vision and Pattern Recognition*, 1999. 73

[88] D. Miyazaki and K. Ikeuchi. Shape estimation of transparent objects by using inverse polarization ray tracing. *IEEE–PAMI*, 29(11):2018–2030, 2007. v, 11, 12, 17, 18, 37, 42

[89] D. Miyazaki, M. Kagesawa, and K. Ikeuchi. Transparent surface modeling by using a pair of polarization images. *IEEE–PAMI*, 26(1):73–82, 2004. 11, 37

[90] Olivier Morel, Christophe Stolz, Fabrice Meriaudeau and Patrick Gorria. Active lighting applied to 3D reconstruction of specular metallic surfaces by polarization imaging. *Applied Optics*, 2006. 11

[91] N. Morris and K. Kutulakos. Reconstructing the surface of inhomogeneous transparent scenes by scatter-trace photography. *IEEE International Conference on Computer Vision*, 2007. v, 11, 12, 14, 17, 18, 37, 41, 70

[92] Nigel J. W. Morris and Kiriakos N. Kutulakos. Dynamic refraction stereo. In *IEEE International Conference on Computer Vision*, pages 1573–1580, Washington, DC, USA, 2005. IEEE Computer Society. ISBN 0-7695-2334-X-02. doi: http://dx.doi.org/10.1109/ICCV.2005.79. 22, 79

[93] N. Morris and K. Kutulakos. Dynamic refraction stereo. *IEEE–PAMI*, 33(8):1518–1531, 2011. 37

[94] H. Murase. Surface shape reconstruction of an undulating transparent object. *IEEE International Conference on Computer Vision*, pp. 313-317, 1990.

[95] H. Murase. Surface shape reconstruction of a nonrigid transparent object using refraction and modulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1045-1052, 1992.

[96] S. G. Narasimhan, S. K. Nayar, B. Sun and S. J. Koppal. Binocular Helmholtz stereopsis. *IEEE International Conference on Computer Vision*, pp. 420-427, 2005.

[97] Srinivasa G. Narasimhan, Mohit Gupta, Craig Donner, Ravi Ramamoorthi, Shree Nayar and Henrik Wann Jensen. Acquiring scattering properties of participating media. *ACM SIGGRAPH*, 2006. v, 14, 16

[98] S. K. Nayar, A. C. Sanderson, L. Weiss and D. Simon. Specular surface inspection using structured highlight and Gaussian images. *IEEE Transactions on Robotics and Automation*, pp. 208-218, 1990.

[99] S.K. Nayar, G. Krishnan, M.D. Grossberg, and R. Raskar. Fast separation of direct and global comefractive index recovery from single-view polarisation images. SIGGRAPH, 2006. vii, 61, 62, 70

[100] S. K. Nayar, X.-S. Fang and T. Boult. Removal of specularities using color and polarization. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 583-590, 1993.

[101] S. K. Nayar Sphereo: determining depth using two specular spheres and a single camera. *SPIE Conference on Optics, Illumination and Image Sensing for Machine Vision*, pp. 245-254, 1988. 7

[102] S. Negahdaripour, H. Sekkati, and H. Pirsiavash. Opitcacoustic stereo imaging, system calibration and 3-d reconstruction. *IEEE Beyond Multiview Geometry*, 2007. 22

[103] David Nistér. An efficient solution to the five-point relative pose problem. *IEEE Pattern Analysis and Machine Intelligence*, 26(6):756-770, June 2004. 77, 82, 83

[104] M. Oren and S.K. Nayar. A theory of specular surface geometry. *IJCV*, 24(2):105–124, 1997.

[105] Margarita Osadchy, David Jacobs and Ravi Ramamoorthi. Using specularities for recognition. *IEEE International Conference on Computer Vision*, pp. 1512-1519, 2003. 20

[106] J. Park and C. Kak. 3D modeling of optically challenging objects. *IEEE Transactions on Visualization and Computer Graphics*, pp. 246-262, 2008.

[107] O. Pizarro, R. Eustice, and H. Singh. Relative pose estimation for instrumented, calibrated imaging platforms. *VIIth Digital Imaging Comp., Tech. and Applications Conf.*, 2003. 22

[108] Helmut Pottmann and Johannes Wallner. *Computational line geometry*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2001. ISBN 3540420584. 26

[109] R. Pless. Using many cameras as one. *IEEE Conference on Computer Vision and Pattern Recognition*, 587-594, 2003. 79

[110] Stefan Rahmann and Nikos Canterakis. Reconstruction of specular surfaces using polarization. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 149-155, 2001.

[111] S. Ramalingam, P. Sturm and S. K. Lodha. Theory and calibration algorithms for axial cameras. *Asian Conference on Computer Vision*, 2006. 79

[112] Srikumar Ramalingam, Peter F. Sturm and Suresh K. Lodha. Towards complete generic camera calibration. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1093-1098, June 2005. 8, 12

[113] Srikumar Ramalingam, Peter Sturm and Suresh Lodha. Theory and Calibration for Axial Cameras. *Asian Conference on Computer Vision*, pp. 704-713, 2006. 80

[114] K. Rinner. Problems of two-medium photogrammetry. *Photogrammetric Engineering*, 35(3):275-282, 1969. 75

[115] S. Roth and M. J. Black. Specular flow and the recovery of surface structure. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1869-1876, 2006. 7, 8

[116] M. Saito, Y. Sato, K. Ikeuchi and H. Kashiwagi. Measurement of surface orientations of transparent objects using polarization in hindlight. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 381-386, 1999. 11

[117] Aswin C. Sankaranarayanan, Ashok Veeraraghavan, Oncel Tuzel and Amit Agrawal. Image invariants for smooth reflective surfaces. *European Conference on Computer Vision*, 2010. 6, 20

[118] Aswin C. Sankaranarayanan, Ashok Veeraraghavan, Oncel Tuzel and Amit Agrawal. Specular surface reconstruction from sparse reflection correspondences. *IEEE Conference on Computer Vision and Pattern Recognition*, 2010. 7, 8

[119] S. Savarese, M. Chen and P. Perona. Local shape from mirror reflections. *International Journal of Computer Vision*, pp. 31-67, 2005.

[120] Y. Y. Schechner, J. Shamir and S. K. Nayar. Polarization-based decorrelation of transparent layers: *IEEE International Conference on Computer Vision*, pp. 814-819, September 1999.

[121] H. Schultz. Retrieving shape information from multiple images of a specular surface. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 195-201, 1994.

[122] Steven Seitz, Yasuyuki Matsushita and Kyros Kutulakos. A theory of inverse light transport. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1440-1447, 2005.

[123] Abd El Rahman Shabayek. Non-central catadioptric sensors auto-calibration. *MSc. Thesis, Erasmus Mundus in Vision and Robotics (VIBOT)*, 2009. 8

[124] M. Shimizu and M. Okutomi. Calibration and rectification for reflection stereo. *IEEE Conference on Computer Vision and Pattern Recognition*, 1-8, June 2008. 79

[125] M. Shortis, E. Harvey and J. Seager. A review of the status and trends in underwater videometric measurement. *SPIE Conference 6491, Videometrics IX*, January 2007. 75

[126] M. R. Shortis and E. S. Harvey. Design and calibration of an underwater stereo-vision system for the monitoring of marine fauna populations. *International Archives of Photogrammetry and Remote Sensing*, 1998. 22

[127] J. D. Smith. The remarkable Ibn al-Haytham. *Mathematical Gazette*, 76(475):189-198, 1992. 8

[128] J. E. Solem, H. Aanaes and A. Heyden. A variational analysis of shape from specularities using sparse data. *International Symposium on 3D Data Processing, Visualization and Transmission*, pp. 26-33, 2004. 17

[129] P. Sturm. Algorithms for plane based pose estimation. *IEEE Conference on Computer Vision and Pattern Recognition*, 706–711, 2000. 61, 67, 69, 73

[130] Peter Sturm and João P. Barreto. General imaging geometry for central catadioptric cameras. *European Conference on Computer Vision*, 2008. 22, 23, 25, 27

[131] P. Sturm and S. Ramalingam. A generic concept for camera calibration. *European Conference on Computer Vision*, 2004. 8, 12, 79

[132] Tomáš Svoboda and Tomáš Pajdla. Epipolar geometry for central catadioptric cameras. *International Journal of Computer Vision*, 49(1):23–37, 2002. ISSN 0920-5691. doi: http://dx.doi.org/10.1023/A:1019869530073. 22

[133] R. Swaminathan, M. Grossberg and S. K. Nayar. A perspective on distortions. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 594-601, 2003.

[134] R. Swaminathn, M. D. Grossberg and S. K. Nayar. Non-single viewpoint catadioptric cameras: geometry and analysis. *International Journal of Computer Vision*, pp. 211-229, 2006. 8

[135] R. Swaminathan, M. Grossberg and S. K. Nayar. Caustics of catadioptric cameras. *IEEE International Conference on Computer Vision*, pp. 2-9, 2001.

[136] Kar-Han Tan, Hong Hua and Narendra Ahuja. Multiview panoramic cameras using mirror pyramids. *IEEE Pattern Analysis and Machine Intelligence*, 26(7), pp. 941-946, July 2004. iv, 6, 7

[137] J.-P. Tardif, P. Sturm, M. Trudeau and S. Roy. Calibration of cameras with radially symmetric distortion. *IEEE Pattern Analysis and Machine Intelligence*, 31(9):1552-1566, 2009. 77, 80

[138] M. Tarini, H. P. A. Lensch, M. Goesele and H.-P. Seidel. 3D acquisition of mirroring objects using striped pattern. *Graphical Models*, 67(4):233-259, 2005.

[139] S. Thirthala and M. Pollefeys. Multi-view geometry of 1D radial cameras and its applications to omni-directional camera calibration. *IEEE International Conference on Computer Vision*, (2):1539-1546, 2005. 80

[140] Y. Tian and S. Narasimhan. Seeing through water: Image restoration using model-based Tracking. *IEEE International Conference on Computer Vision*, 2009. 41

[141] Tali Treibitz, Yoav Y. Schechner, and Hanumant Singh. Flat refractive geometry. *IEEE Conference on Computer Vision and Pattern Recognition*, 0:1–8, 2008. doi: http://doi.ieeecomputersociety.org/10.1109/CVPR.2008.4587844. 22, 23, 76, 79

[142] Shinji Umeyama and Guy Godin. Separation of diffuse and specular components of surface reflection by use of polarization and statistical analysis of images. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 26(5):639-647, 2004.

[143] Y. Vasilyev, Y. Adato, T. Zickler and O. Ben-Shahar. Dense specular shape from multiple specular flows. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2008. 8

[144] J. Wang and K. J. Dana. Relief texture from specularities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 446-457, 2006.

[145] R. J. Woodman. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, pp. 139-144, 1980. 12

[146] Masaki Yamazaki, Sho Iwata and Gang Xu. Dense 3D reconstruction of specular and transparent objects using stereo cameras and phase-shift method. *Asian Conference on Computer Vision*, pp. 570-579, 2007.

[147] R. Yang, M. Pollefeys and G. Welch. Dealing with textureless regions and specular highlights - A progressive space carving scheme using novel photoconsistency measure. *IEEE International Conference on Computer Vision*, pp. 576-584, 2003.

[148] S.-K. Yeung, T.-P. Wu, C.-K. Tang, T.F. Chan, and S. Osher. Adequate reconstruction of transparent objects on a shoestring budget. *IEEE Conference on Computer Vision and Pattern Recognition*, 2011. 12, 41

[149] Jingyi Yu and L. McMillan. General linear cameras. *European Conference on Computer Vision*, 2004.

[150] Jingyi Yu and L. McMillan. Modelling reflections via multiperspective imaging. *IEEE Conference on Computer Vision and Pattern Recognition*, June 2005.

[151] J. Y. Zheng, Y. Fukugawa and N. Abe. 3D surface estimation and model construction from specular motion in image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 513-520, 1997.

[152] J. Y. Zheng and A. Murata. Acquiring 3D object models from specular motion using circular lights illumination.

[153] T. Zickler, P. N. Belhumeur and D. Kriegman. Helmholtz steropsis: Exploiting reciprocity for surface reconstruction. *International Journal of Computer Vision*, pp. 215-227, 2002.

[154] A. Zisserman, P. Gilbin and Andrew Blake. The information available to a moving observer from specularities. *Image and Vision Computing*, pp. 38-42, 1989. *IEEE Internation Conference on Computer Vision*, pp. 1101-1108, 1998. 8

[155] Todd Zickler, Peter N. Belhumeur and David J. Kriegman. Helmholtz Stereopsis: exploiting reciprocity for surface reconstruction. *European Conference on Computer Vision*, 2002. 12

2033

Appendices

.1 Partial Polarization

When unpolarized light with electric field magnitude || falls on a dielectric transparent object at an angle θ_1 and is refracted at an angle θ_2, the reflected and transmitted portions of light are partially polarized. Given refractive indices n_1 and n_2, let us compute the perpendicular and parallel components of the reflected ray. A similar derivation can be obtained for the transmitted ray.

$$\|\mathbf{E}^1\|\cos(\psi) = k\sqrt{R_s}\|\mathbf{E}\|\cos(\phi) \quad (1)$$

$$\|\mathbf{E}^1\|\sin(\psi) = k\sqrt{R_p}\|\mathbf{E}\|\sin(\phi) \quad (2)$$

$$\tan(\psi) = \sqrt{\frac{R_p}{R_s}}\tan(\phi) \quad (3)$$

$$\psi = \tan^{-1}(\sqrt{\frac{R_p}{R_s}}\tan(\phi)) = g(\phi) \quad (4)$$

where k is the normalizing factor to make sure both sides are components of unit vectors, ϕ denotes the angle between the electric field and its parallel component, and ψ denotes the same angle for the reflected ray. Since the incoming light is unpolarized, the probability of any ϕ occuring is $\rho(\phi) = \frac{1}{\pi}$.

Thus denoting $\lambda = \sqrt{\frac{R_p}{R_s}}$, the probability of ψ can be computed using a variable transformation.

$$\rho(\psi) = \frac{1}{|g'(\phi)|}\rho(\phi) \tag{5}$$

$$g'(\phi) = \frac{1}{1+\lambda^2\tan^2(\phi)}(\lambda\sec^2(\phi)) \tag{6}$$

$$= \frac{1}{1+\lambda^2\tan^2(\phi)}(\lambda(1+\tan^2(\phi))) \tag{7}$$

$$= \frac{1}{1+\tan^2(\psi)}\frac{\lambda^2+\tan^2(\psi)}{\lambda}(\text{see } 3) \tag{8}$$

$$= \frac{\lambda^2+\tan^2(\psi)}{\lambda(1+\tan^2(\psi))} = \frac{\lambda^2-1+\sec^2(\psi)}{\lambda\sec^2(\psi)} \tag{9}$$

$$\frac{1}{|g'(\psi)|} = \frac{\lambda\sec^2(\psi)}{|\lambda^2-1+\sec^2(\psi)|} \tag{10}$$

$$= \frac{\lambda}{|(\lambda^2-1)\cos^2(\psi)+1|} \tag{11}$$

This Finally, this gives $\rho(\psi)$ in the case of both reflection and transmission, as

$$\rho(\psi)_l = \frac{\sqrt{R_sR_p}}{|(R_p-R_s)\cos^2(\psi)+R_s|}\frac{1}{\pi} \tag{12}$$

$$\rho(\psi)_r = \frac{\sqrt{T_sT_p}}{|(T_p-T_s)\cos^2(\psi)+T_s|}\frac{1}{\pi} \tag{13}$$

Notice that $R_p = 0$ at *Brewster's angle*, which makes $\rho(\psi) = 0$ for all angles except $\psi = 0$ (where denominator reduces to R_p, and so $\rho(0)$ is undefined). This agrees with the theory that reflected light is linearly polarized perpendicular to the plane of refraction. When this reflected or transmitted light reaches a second refractive surface, the Fresnel equations apply on the basis of a new plane of refraction which is different from the previous one. Thus the new electric field $\mathbf{E}^2$ is a rotated version of the electric field associated with the reflected or transmitted light $\mathbf{E}^1$. Fortunately, this just results in a shift in angle space, so $\psi' = \psi - \alpha$ where α is the angle between the two planes of refractions. This follows immediately from the fact that the two planes of refraction contain the transmitted or reflected ray, and hence thier normals must lie in a plane perpendicular to the direction of propagation

of the ray.

.2 Second Bounce Radiance Ratio

Following equation 3.13, we now determine the radiance ratio for the case when two reflections occur. The other cases for any combination of two bounces follows the same pattern. In order to compute the radiance ratio after the first bounce, first we need to compute the intensity of a single ray after the first bounce.

$$^1\|\mathbf{E}\| = \sqrt{^1R_s E^2 \cos(\phi) + {}^1R_p E^2 \sin(\phi)} \tag{14}$$

Since $\tan(\phi) = \sqrt{\frac{^1R_s}{^1R_p}}\tan(\psi)$, we can modify the above equation by noting

$$\cos^2(\phi) = \frac{^1R_p \cos^2(\psi)}{^1R_p \cos^2(\psi) + {}^1R_s \sin^2(\psi)} \tag{15}$$

$$\sin^2(\phi) = \frac{^1R_s \sin^2(\psi)}{^1R_p \cos^2(\psi) + {}^1R_s \sin^2(\psi)} \tag{16}$$

Thus, we get

$$^1\|\mathbf{E}\| = {}^1E = E\sqrt{\frac{^1R_s\, ^1R_p}{^1R_p \cos^2(\psi) + {}^1R_s \sin^2(\psi)}} \tag{17}$$

$$^1E^2 = E^2 \frac{^1R_p}{k^2 \cos^2(\psi) + \sin^2(\psi)} \tag{18}$$

Substituting ${}^1E^2$ for E^2 in equation 3.13, we get the intensity for second bounce as

$$
\begin{aligned}
{}^2I_{ll} &= {}^1E^2 \int_l^u ({}^2R_s \cos^2(\psi) + {}^2R_p \sin^2(\psi))\rho(\psi)d\psi && (19)\\
&= E^2\ {}^1R_p\ {}^2R_p \int_l^u \frac{\cos^2(\psi)l^2 + \sin^2(\psi)}{k^2\cos^2(\psi) + \sin^2(\psi)}\rho(\psi)d\psi && (20)\\
&= E^2\ {}^1R_p\ {}^2R_p \frac{1}{2}\left(\frac{({}^2R_s\ {}^1R_p + {}^1R_s\ {}^2R_p)\sin^2(\alpha) + ({}^1R_s\ {}^2R_s + {}^1R_p\ {}^2R_p)\cos^2(\alpha)}{{}^1R_p\ {}^2R_p}\right) && (21)\\
&= E^2\frac{1}{2}\left(({}^2R_s\ {}^1R_p + {}^1R_s\ {}^2R_p)\sin^2(\alpha) + ({}^1R_s\ {}^2R_s + {}^1R_p\ {}^2R_p)\cos^2(\alpha)\right) && (22)
\end{aligned}
$$

where $k = \sqrt{\frac{{}^1R_p}{{}^1R_s}}$ and $l = \sqrt{\frac{{}^2R_s}{{}^2R_p}}$

Coordinate transformation After a light ray passes through the first bounce, the angle its electric field makes with the ${}^1\pi$ is ψ. However, when it strikes the second surface in plane of reflection/refraction ${}^2\pi$, the angle made with ${}^2\pi$ is no longer ψ. Since ${}^1\pi$ and ${}^2\pi$ are normal vectors perpendicular to the middle segment of the light path, they lie in the plane that contains the electric field ${}^2\mathbf{E}$. Thus, this ray makes an angle $\psi - \alpha$ with ${}^2\pi$, where α is the angle between ${}^1\pi$ and ${}^2\pi$.