A Theory of Refractive and Specular 3D Shape by Light-Path Triangulation

KIRIAKOS N. KUTULAKOS* AND ERON STEGER

Department of Computer Science, University of Toronto, Toronto ON, Canada M5S 3G4

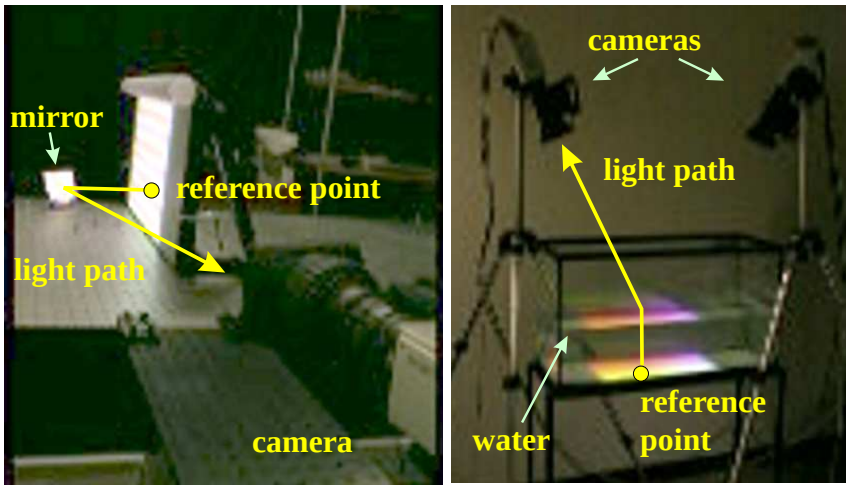
kyros@cs.toronto.edu

esteger@cs.toronto.edu

Abstract. We investigate the feasibility of reconstructing an arbitrarily-shaped specular scene (refractive or mirror-like) from one or more viewpoints. By reducing shape recovery to the problem of reconstructing individual 3D light paths that cross the image plane, we obtain three key results. First, we show how to compute the depth map of a specular scene from a single viewpoint, when the scene redirects incoming light just once. Second, for scenes where incoming light undergoes *two* refractions or reflections, we show that three viewpoints are sufficient to enable reconstruction in the general case. Third, we show that it is impossible to reconstruct individual light paths when light is redirected more than twice. Our analysis assumes that, for every point on the image plane, we know at least one 3D point on its light path. This leads to reconstruction algorithms that rely on an "environment matting" procedure to establish pixel-to-point correspondences along a light path. Preliminary results for a variety of scenes (mirror, glass, etc) are also presented.

Keywords: 3D photography, 3D scanning, shape acquisition, multi-media photogrammetry, reflectance modeling, shape from refraction, shape from specularities, triangulation, stereo, transparency, specular reflection

*Part of this research was conducted while K. Kutulakos was serving as a Visiting Scholar at Microsoft Research Asia.



(a) (b)

Figure 1: Viewing a known reference point indirectly via (a) an opaque specular scene (a mirror) and (b) a transparent specular scene (a volume of water).

1 Introduction

The reconstruction of general specular scenes, either refractive or mirror-like, is one of the few remaining open problems in visual reconstruction. Examples include scenes that contain glass objects, mirrors, or liquids, where refraction and specular reflection dominate the image formation process. Such scenes cannot be reconstructed by laser scanners or by 3D reconstruction algorithms designed for objects that scatter incident light (e.g., [1–3]). Reconstructing such scenes, on the other hand, could have implications in many disciplines, including graphics [4, 5], optics [6, 7], 3D scanning [8, 9], and fluid modeling [10].

Specular objects do not have an "appearance" of their own—they simply distort the appearance of other objects nearby, creating an *indirect view* of the original objects. Unlike perspective images, where 3D points project along straight lines, indirect views are created by light that travels along a piecewise-linear light path (Figure 1). The complexity of this projection process and the difficulty of inverting it has brought about new image-based techniques, such as environment matting [4, 5, 11], that side-step 3D reconstruction altogether. Instead of computing shape, they compute the shape's effect on appearance—all they recover is a function that maps points on a pattern placed near the scene to pixels in the pattern's distorted, indirect view.

In this paper, we investigate the reconstruction of such scenes with an approach that seeks to invert the indirect projection process. Despite the problem's apparent intractability in the general

case, it is possible to characterize the class of reconstructible scenes and to develop simple reconstruction algorithms for some important cases. In particular, our work considers three questions:

- suppose we are given a function that maps each point in the image to a 3D "reference point" that indirectly projects to it; can we recover the point's light path?
- if so, under what conditions?
- how do we design reconstruction algorithms that do not impose any *a priori* constraints on the shape of the unknown specular scene?

Little is known about how to address these questions in the general case, although specialized reconstruction algorithms for a few cases have been developed. The earliest algorithms come from *multi-media photogrammetry* [12, 13], where the scene is assumed to have a known parametric form. These approaches solve a generalized structure-from-motion problem that takes into account refractions and reflections caused by parametric surfaces with a few known degrees of freedom (e.g., underwater objects viewed from above a planar sea surface). An algorithm along these lines was recently proposed by Ben Ezra and Nayar [14] for reconstructing glass objects modeled as super-ellipsoids. Knowledge of a scene's low-order parametric form implies that these techniques cannot be used for reconstructing objects with fine detail or with a complicated, unknown shape.

Most computer vision research on the topic has followed a "shape-from-distortion" approach for reconstructing either mirrors [9, 15] or liquids [16–18]. In this approach, 3D shape is recovered by analyzing the distortion of a known pattern placed near the specular surface. Unfortunately it is impossible, in general, to reconstruct the 3D shape of an unknown specular scene from just one image of a nearby pattern. This has prompted a variety of assumptions, including approximate planarity [17–19], surface smoothness [15], integrability [9], far-field illumination [27] and special optics [10, 16, 20]. These approaches are restricted to the simplest forms of indirect viewing, where light bounces at most once before reaching the camera (e.g., by reflecting off a mirror or refracting once through the air-water boundary). Moreover, specialized research on reconstructing specular transparent objects has followed one of three basic approaches—they either ignore the object's specular properties, relying exclusively on the object's silhouette for reconstruction [4], they analyze the polarization of light specularly reflected from its surface [21], or they reduce reconstruction to a standard computerized tomography problem [22, 23]. Unfortunately, silhouette-based approaches are limited to recovering a visual hull approximation, and polarization-based analysis is difficult when transmission, rather than specular reflection, dominates image formation. While computerized tomography enables reconstruction of very complex semi-transparent shapes when light propagation is linear [23, 24], enforcing linear propagation through glass media is fairly

intrusive (e.g., it requires immersing the glass object in a semi-transparent liquid with an identical refraction index [22]) and may even be impossible to achieve (e.g., when the object contains inaccessible cavities or opaque regions).

Our goal is to develop a general framework for analyzing specular scenes that does not impose *a priori* assumptions on the shape of their surfaces or the nature of their media (e.g., opaque or transparent). To achieve this, we formulate the reconstruction of individual light paths as a geometric constraint satisfaction problem that generalizes the familiar notion of triangulation to the case of indirect projection.

Our approach can be thought of as complementing two lines of recent work. Research on *environment matting and generalized imaging models* [5, 25, 26] represents an arrangement of cameras, mirrors and lenses as an abstract function that maps 3D points or 3D rays to points on the image plane. These techniques focus on computing this function and treat the arrangement itself as an unknown "black box." In contrast, here we assume that this function is known and study the problem of reconstructing the arrangement. Work on *specular stereo* [27–30] relies on a two-camera configuration or a moving observer to reconstruct a mirror-like object. These algorithms solve the light path reconstruction problem for one specific case; our framework leads to several generalizations, including a stronger two-view result [31] that enables reconstruction of a refractive scene even when its refractive index is unknown.

On the theoretical side, our work has five key contributions. First, we provide a unified analysis of refractive and mirror-like scenes, leading to algorithms that work for both problems. Second, we characterize the set of reconstructible scenes in a way that depends only on the number of vertices along a light path. As such, our results apply to any specific scene geometry that produces paths of a given length. Third, we identify a very simple algorithm for computing the depth map of a mirror surface from one viewpoint. The algorithm relies on knowledge of a function that maps each image point to two known reference points along its light path and places no restrictions on shape, except that light must bounce exactly once before reaching the camera. Fourth, we establish the most general class of scenes that can be reconstructed using an efficient, stereo-like algorithm: these are scenes where light bounces twice before reaching the camera. To our knowledge, this problem, which requires three viewpoints to solve it, has not been previously analyzed. Fifth, we show that, while efficient algorithms may not exist for scenes with light paths of length $K \geq 3$, there is enough information in $3(K-1)$ viewpoints to reduce shape ambiguities to a discrete set.

Even though our emphasis here is on the underlying theory, we present preliminary results on real scenes, both refractive and mirror-like. These results have several implications. First,

they show that we can reconstruct mirror surfaces with a technique whose accuracy is bounded by the calibration accuracy of a single stationary camera and by the accuracy of environment matting (which can be very high using well-known techniques [5, 32]). Second, it is possible to reconstruct each point on a specular 3D scene (mirror, liquid, glass) independently of all other points. This allows reconstruction of scenes with fine surface detail and/or discontinuities. Third, we can compute a separate depth *and* a separate normal for each surface point; this is unlike typical stereo or laser-scanning techniques (which compute a point-set that must be differentiated to get normals) or photometric stereo (which computes a normal map that must be integrated to obtain depth). As such, our algorithms yield richer 3D data for inferring an object's unknown shape [3, 33].

2 Light-Path Triangulation

Perspective projection requires that every 3D point projects to an image along a straight line. When the scene is composed of refractive or mirror-like objects, this linear projection model is not valid anymore. Here we extend this model by studying *indirect* projections of 3D points. Informally, indirect projection occurs anytime a point is viewed indirectly, via one or more specular surfaces.

Consider a scene that is viewed from one or more known viewpoints and contains one or more objects of unknown shape. We assume that each object is a volume composed of a homogeneous medium (opaque or transparent) and whose surface is smooth, i.e., it does not contain surface irregularities that scatter the incident light. In this case, the propagation of light through the scene is characterized by three basic processes [34, 35]—*specular reflection* at an object's surface, *specular transmission (i.e., refraction)* at the surface of a transparent object, and *linear propagation* within an object's interior and through empty space.

Given an arbitrary 3D point $\mathbf{p}$, a known viewpoint $\mathbf{c}$, and a known image plane, the point's projection is determined by the 3D path(s) that light would trace in order to reach that viewpoint (Figure 2). We use the term *light path* to refer to such a path. If a light path exists, it will be a piecewise-linear curve between $\mathbf{p}$ and $\mathbf{c}$ whose vertices, if any, will always lie on the surface of some object in the scene. The number of vertices along a path is therefore equal to the number of surfaces it intersects. In general, there may be more than one light path connecting a 3D point to a viewpoint, or there may be none at all.[1] We say that point $\mathbf{q}$ is an *indirect projection* of $\mathbf{p}$ if there is a light path between $\mathbf{p}$ and $\mathbf{c}$ that crosses the image plane at $\mathbf{q}$.

[1] See [36] for a camera-mirror arrangement that forces scene points to indirectly project twice onto the image plane.

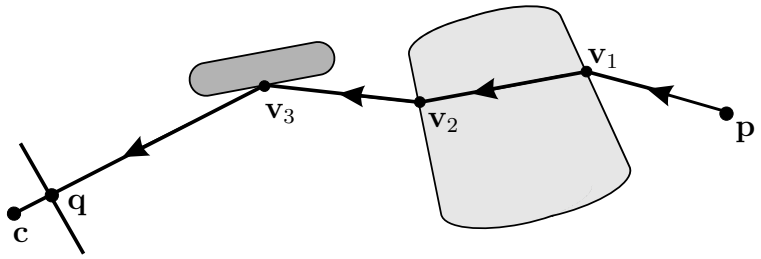


Figure 2: An example light path. The dark gray region denotes a mirror-like object and the light gray region a transparent object. Here, the light path from $\mathbf{p}$ intersects three surfaces before reaching point $\mathbf{q}$ on the image plane, and therefore has three vertices, $\mathbf{v}_1, \mathbf{v}_2$ and $\mathbf{v}_3$, and four rays. In light-path triangulation, the coordinates of $\mathbf{c}, \mathbf{q}$ and $\mathbf{p}$ are known and the goal is to determine the coordinates and normals of the vertices. By convention, we order vertices and rays along a path according to the direction of light travel.

2.1 The Light-Path Triangulation Problem

Suppose the specular scene is viewed from N known viewpoints. We assume that for every point on the associated image planes there is a unique light path that describes light propagation toward that point.[2] Furthermore, suppose we are given a function which tells us, for every such point, the 3D coordinates of M "reference points" that project to that point indirectly (Figure 3). Now, suppose we choose a point $\mathbf{q}$ on one of the image planes and assign it a "depth" value, i.e., a hypothetical distance to the last vertex along its light path. Under what conditions can we decide unambiguously the correctness of this depth? Our goal is to answer this question in the general case, i.e., for smooth scenes of arbitrary shape, $N \geq 1$ viewpoints, $M \geq 1$ known reference points, and light paths with $K \geq 1$ vertices. To simplify our exposition, we assume without loss of generality that all light paths have the same number, K, of vertices and that this number is known.

When we assign a depth d to a point on the image plane, we define the 3D position of one specular point, $\mathbf{v}^d$, along the ray through the selected image point. If that depth is correct, $\mathbf{v}^d$ would redirect light toward all N viewpoints in a way that is consistent with the laws of refraction and reflection, as well as the known function that maps image points to reference points. Specifically, light would travel along N distinct light paths whose last vertex is $\mathbf{v}^d$ (Figure 3). These paths define a graph, that we call the *light network* for depth d. The network connects the N perspective projections of $\mathbf{v}^d$ to their corresponding reference points.

[2] More generally, our theory applies when the mapping from image points to light paths is one-to-L with L finite and bounded; for simplicity of presentation, however, we assume $L = 1$ in this paper.

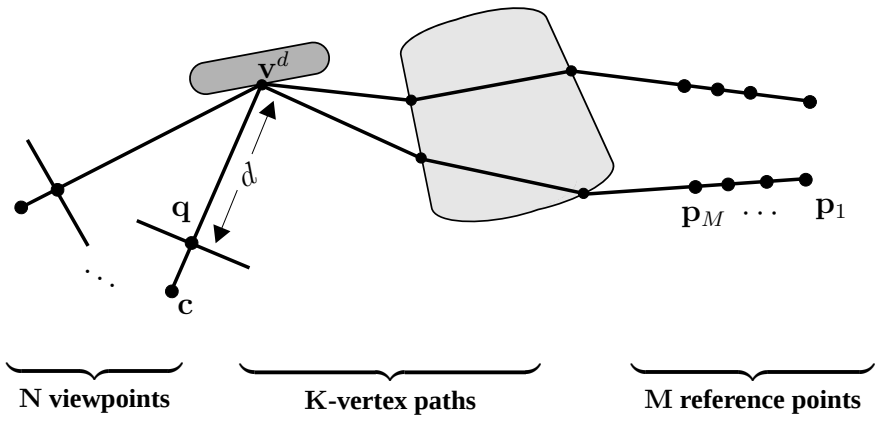


Figure 3: Basic geometry of $\langle N, K, M\rangle$-triangulation.

Definition 1 (Consistent Light Network) *The light network for depth d is* consistent *if we can assign a normal to $\mathbf{v}^d$ and 3D coordinates and normals to its other vertices so that the resulting light paths are consistent with the laws of reflection and refraction.*

Definition 2 ($\langle\mathbf{N}, \mathbf{K}, \mathbf{M}\rangle$-Triangulation) *Assigns a depth d to a given image point so that the resulting light network is consistent.*

Definition 3 (Tractability) *A triangulation problem is* tractable *for a given image point if its solution space is a 0-dimensional manifold, i.e., it is a collection of isolated depth values.*

Intuitively, the minimum M and N needed to make triangulation tractable for a given path length K indicate the problem's intrinsic difficulty. We use the term *light-path triangulation* to refer to the entire family of $\langle N, K, M\rangle$-triangulation problems.

Light-path triangulation differs from traditional stereo triangulation in three important ways. First, unlike stereo where at least two viewpoints are needed for reconstruction, tractable light-path triangulation is possible even with just one viewpoint (Section 3.1). Second, unlike stereo where a single point is reconstructed from a pair of intersecting 3D rays, here we must reconstruct the 3D coordinates of *all* $N(K-1)+1$ points in a light network, to guarantee consistency. Third, while stereo triangulation does not provide surface normal information, light-path triangulation reconstructs normals as well. Hence, even though it is harder to solve, light-path triangulation yields richer scene descriptions than stereo both in terms of density (i.e., number of reconstructed points) and content (i.e., points and normals).

2.2 Basic Properties of a Light Path

In principle, it is always possible to express a light-path triangulation problem as a system of non-linear equations that govern light propagation through the scene. Rather than study the analytical form of those equations, which can be quite complex, we take a geometric approach. In particular, we express $\langle N, K, M \rangle$-triangulation as a geometric constraint satisfaction problem whose solution space depends on just three properties (Figure 4):

- **Planarity Property:** Light propagation at a vertex always occurs on a single plane that contains the surface normal. That is, the vectors $\mathbf{n}, \mathbf{d}^{\text{in}}$ and $\mathbf{d}^{\text{out}}$ are always coplanar.
- **Deflection Property:** If we know the refractive index and know any two of vectors $\mathbf{n}, \mathbf{d}^{\text{in}}, \mathbf{d}^{\text{out}}$, we can determine uniquely the third vector. Moreover, this relation is a local diffeomorphism.[3]
- **Double-Correspondence Property:** If we are given two distinct reference points that project indirectly to the same image point, the first ray on the image point's light path must be the line that passes through both reference points.

Note that all three properties hold for reflected and for refracted light. As a result, our analysis does not distinguish between these two different types of light propagation, making our theoretical results applicable to scenes with mirror-like or refractive objects, or both.[4]

While not previously used for reconstruction, the Double-Correspondence Property has been noted in the context of environment matting [5] and camera calibration [25]. Here, it highlights a fundamental difference between light-path triangulations where two or more reference points are known per image point ($M \geq 2$) versus just one ($M = 1$): two or more reference points provide complete information about the 3D ray along which light propagates *before* it enters the scene, which is impossible to get from just one reference point.[5] This distinction is especially important in interpreting the results of our analysis.

[3] Recall that a smooth map, f, between two manifolds is a local diffeomorphism at a point p if its derivative, $\mathrm{d}f_p$, is one-to-one and onto [37].

[4] In fact, our analysis covers the case where a scene point causes a refraction along the light path of a point in an image and causes a specular reflection along the path of some other image point (in the same or another viewpoint).

[5] Note that the *direction* of this 3D ray can be determined from a single reference point "at infinity," i.e., located far from the specular scene [27]. Since this direction is not sufficient to localize the ray in 3D, the information provided by a single reference point at infinity is weaker than knowledge of two reference points.

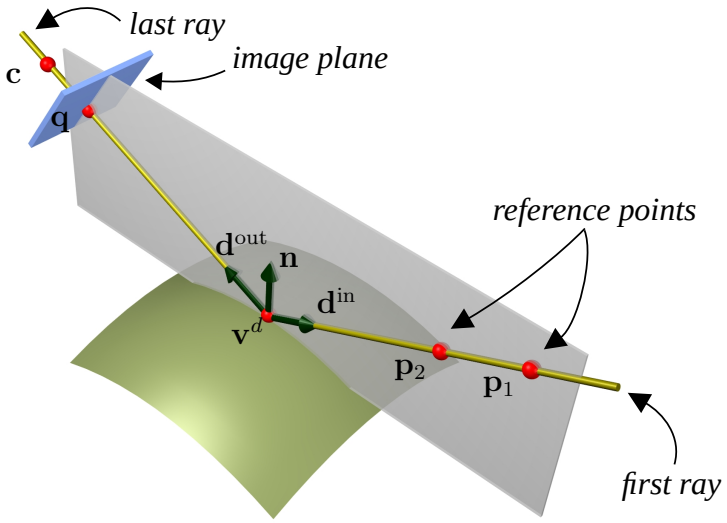


Figure 4: Visualizing the three properties of a light path. Vectors $\mathbf{n}, \mathbf{d}^{\text{in}}, \mathbf{d}^{\text{out}}$ are always coplanar. In specular reflection, shown above, the angle between $\mathbf{n}$ and $\mathbf{d}^{\text{in}}$ is always equal to that of $\mathbf{n}$ and $\mathbf{d}^{\text{out}}$. In specular transmission, Snell's law states that the ratio of sines of these angles is equal to the relative index of refraction [34]. Hence, knowing one angle allows us to determine the other in both cases.

3 Tractable Light-Path Triangulations

Our main theoretical result is an enumeration of all tractable light-path triangulation problems (Figures 5, 6):

Theorem 1 *The only tractable $\langle N, K, M\rangle$-triangulations are shown in the tables below:*

One reference point $(\mathbf{M = 1})$			
	$K = 1$	$K = 2$	$K \geq 3$
$N = 1$			
$N \geq 2$	✓ ×		

Two or more reference points $(\mathbf{M \geq 2})$			
	$K = 1$	$K = 2$	$K \geq 3$
$N = 1$	✓ ×		
$N = 2$	✓ ×		
$N = 3$	✓ ×	✓	
$N \geq 4$	✓ ×	✓ ×	

where '✓' marks tractable problems where the scene is either known to be a mirror or its refractive index is known; '×' marks tractable problems where the refractive index (or whether it is a mirror) is unknown; and blanks correspond to intractable cases.

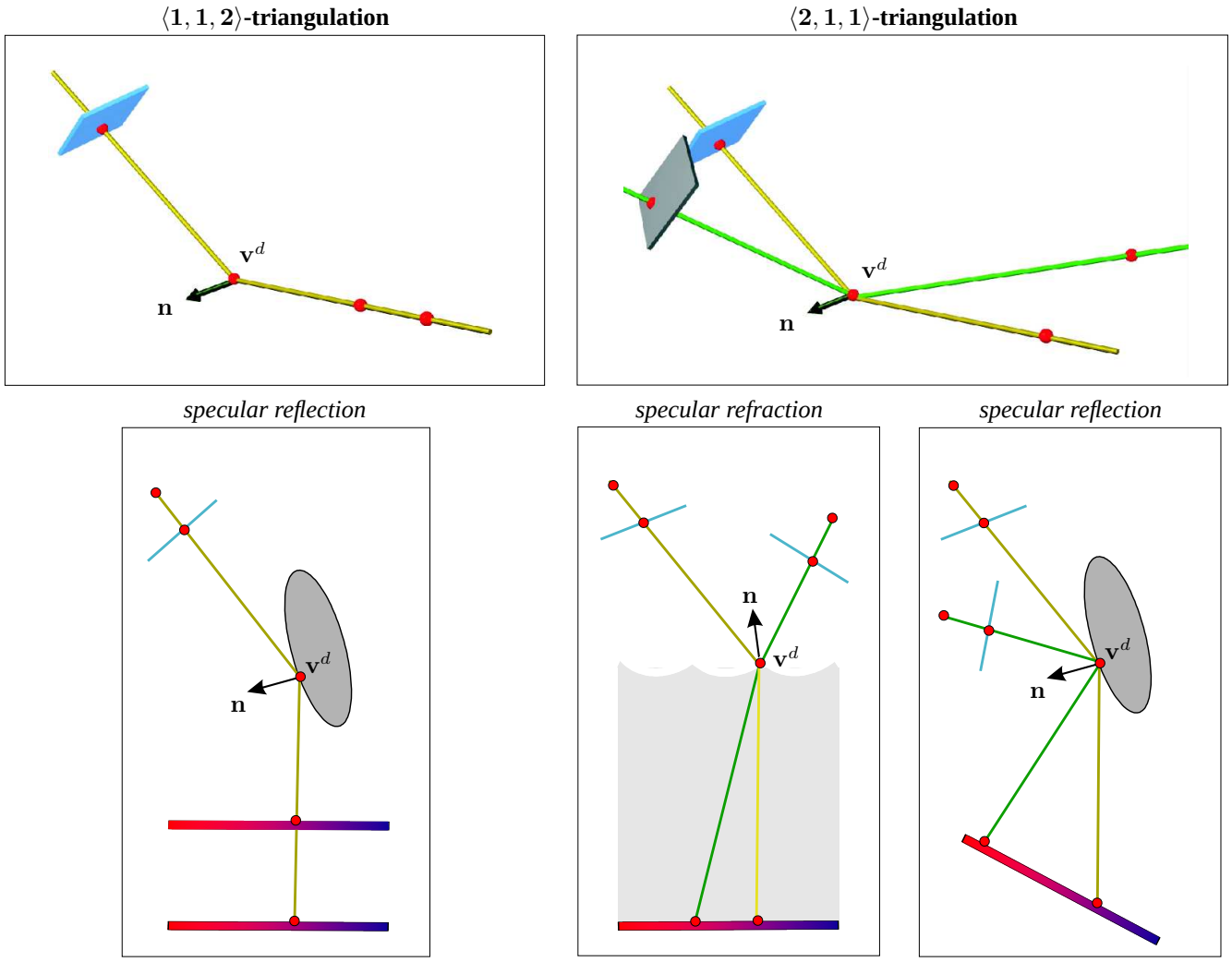


Figure 5: Two of the basic tractable light-path triangulation problems. The third tractable problem is shown in Figure 6. *Top row:* General 3D geometry of light paths and normals. Similarly-colored rays are on the same light path. Unknown vertices and normals are indicated along each path (there is only one unknown vertex and normal for these two problems). *Bottom row:* Top-down view of the unknown scenes and light paths, and the known reference points, for specific instances of these problems. In practice, the known reference points lie on a movable LCD panel whose 3D position is always known (indicated by the red-blue color gradients). Two instances are shown for the case of $\langle 2, 1, 1 \rangle$-triangulation, corresponding to a transparent and a mirror-like scene, respectively.

We obtain this result through a case-by-case analysis in which the three properties of Section 2.2 are applied to the above cases. Proofs for the cases of $\langle 1, 1, 2 \rangle$-triangulation and $\langle 3, 2, 2 \rangle$-triangulation are given in Sections 3.1 and 3.2, respectively. Each of these proofs is constructive

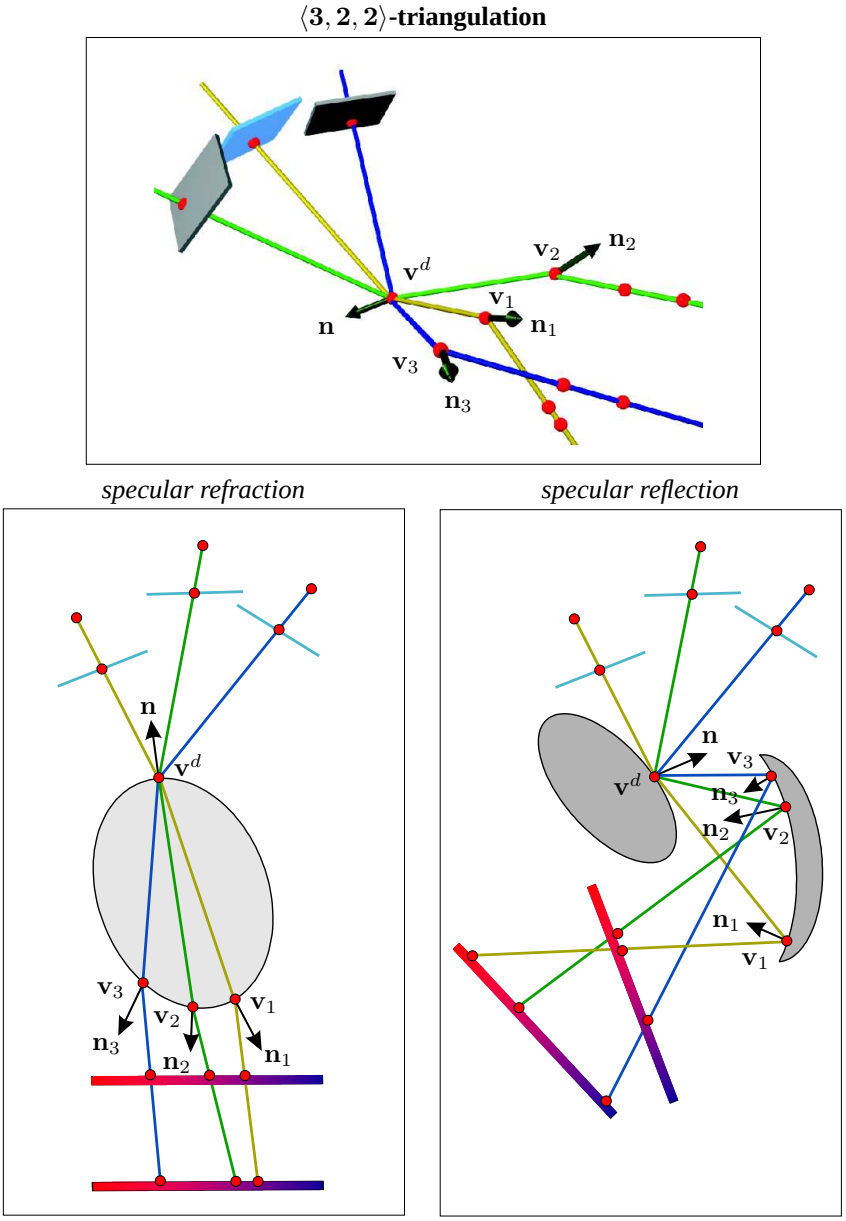


Figure 6: $\langle 3, 2, 2\rangle$-triangulation. We use the same conventions as in Figure 5. Note that in addition to reconstructing transparent objects, $\langle 3, 2, 2\rangle$-triangulation enables reconstruction of mirror-like scenes from the "reflection of the reflection" of the reference points. The instances shown above are not exaustive; our theory also covers cases where some of the light paths through $\mathbf{v}^d$ are caused by refraction and some by reflection.

and leads directly to a reconstruction algorithm. See [31] for a detailed investigation of a third case, $\langle 2, 1, 1\rangle$-triangulation, which includes a proof, algorithmic details, and experimental results

on reconstructing dynamic surfaces of liquids.

Theorem 1 can be interpreted both as a negative and as a positive result. On the negative side, it tells us that light-path triangulation quickly becomes intractable for scenes where a light path intersects many surfaces. Moreover, our capabilities are severely limited when $M = 1$, i.e., when one known reference point projects to each image point.

On the positive side, the theorem identifies three non-trivial cases that *are* tractable: (1) reconstructing a mirror from just one viewpoint; (2) reconstructing a refractive surface with an unknown refractive index from two viewpoints; and (3) using three viewpoints to reconstruct scenes that refract or reflect light twice.

Theorem 1 also highlights a fundamental asymmetry between the number of known reference points (M) and the number of viewpoints (N). Intuitively, the information we obtain by indirectly viewing two known reference points per image point cannot be replaced by viewing just one reference point and increasing the number of viewpoints.

3.1 Mirrors: One Viewpoint, Two Reference Points

The tractability of $\langle 1, 1, 2\rangle$-triangulation is a trivial consequence of the geometry of Figure 4. This is because knowledge of two distinct reference points on a one-vertex light path means that we know both rays on that path. The depth of the path's vertex is therefore given by the intersection of these two rays.

Specifically, suppose that we know the two reference points, $\mathbf{p}_1, \mathbf{p}_2$, that indirectly project to image point $\mathbf{q}$, and suppose that we do not know the scene's refractive index or whether it is a mirror. In this case, the first ray along the light path is the ray through points $\mathbf{p}_1$ and $\mathbf{p}_2$, and the second ray is the ray through $\mathbf{q}$ and the camera's known viewpoint, $\mathbf{c}$. The unique depth solution is given by

$$d = \frac{\|(\mathbf{p}_1 - \mathbf{c}) \times \mathbf{d}^{\text{in}}\|}{\|\mathbf{d}^{\text{out}} \times \mathbf{d}^{\text{in}}\|} \tag{1}$$

where $\mathbf{d}^{\text{in}}$ and $\mathbf{d}^{\text{out}}$ are the unit vectors in the direction of the path's two rays. Note that if we also know that $\mathbf{q}$'s light path is caused by specular reflection, the surface normal at the path's vertex is uniquely determined—it is simply the unit vector in the direction of the bisector, $(\mathbf{d}^{\text{in}} + \mathbf{d}^{\text{out}})/2$.[6]

While $\langle 1, 1, 2\rangle$-triangulation has a very simple solution, we are not aware of prior work that uses

[6]When this information is not available, one additional viewpoint is sufficient to determine both the normal and the scene's specular properties (i.e., whether it is reflective or refractive, and the refractive index).

it for reconstructing specular scenes.[7]

3.2 Glass: Three Viewpoints, Two Reference Points

Figure 7a shows the geometry of a typical light path in the case of $\langle 3, 2, 2\rangle$-triangulation. The path will contain two vertices and three non-coplanar rays, two of which are known (i.e., the first and the last ray). To determine the path uniquely we therefore need just two additional scalars—the depth d of its second vertex and the position, δ, of its first vertex along the first ray. To prove the tractability of $\langle 3, 2, 2\rangle$-triangulation for a known refractive index we show that, in general, only an isolated set of (d, δ)-pairs will define a consistent light network. Intuitively, this is because when two additional viewpoints are available, almost every (d, δ)-pair will produce at least one "invalid" light path, i.e., a path whose rays do not meet in 3D for one of those viewpoints (e.g., the green and blue paths in Figures 7b and 7c, respectively).

Proposition 1 *(a)* $\langle 3, 2, 2\rangle$*-triangulation is tractable for almost all points on a generic surface with known refractive index. (b)* $\langle 3, 2, 2\rangle$*-triangulation is intractable when the refractive index is unknown.*

Beyond showing that it is possible to reconstruct general doubly-refracting and doubly-reflecting scenes, our analysis suggests a reconstruction algorithm: it tells us that we can reconstruct all four vertices and normals in the light network of a pixel (Figure 5) by conducting a 2D search in (d, δ)-space. The search is for a pair (d, δ) that produces valid light paths in all three views.

Geometrically, the proof of Proposition 1 can be thought of as exploiting two basic observations: (1) the set of all depth and normal assignments consistent with a single viewpoint forms a 2D "constraint surface" in $\mathbb{R} \times \mathbb{S}^2$; and (2) the common intersection of three such surfaces (i.e., one for each viewpoint) will in general be a set of isolated points. In the following, we develop a constructive proof that formalizes these intuitions.

Proof of Proposition 1(a): For concreteness, assume that the "true" light path of every image point contains two refractive vertices (Figure 6, bottom left). Paths where one or both of their vertices are reflective can be treated in an identical way.

[7] The Double-Correspondence Property was used in [25] to recover the caustic of a mirror-based imaging system. This caustic, however, does not coincide with the mirror's surface and, hence, their technique is not equivalent to $\langle 1, 1, 2\rangle$-triangulation. More recently, and independently from our own work [38], Bonfort, Sturm and Gargallo [39] reported an algorithm identical to $\langle 1, 1, 2\rangle$-triangulation.

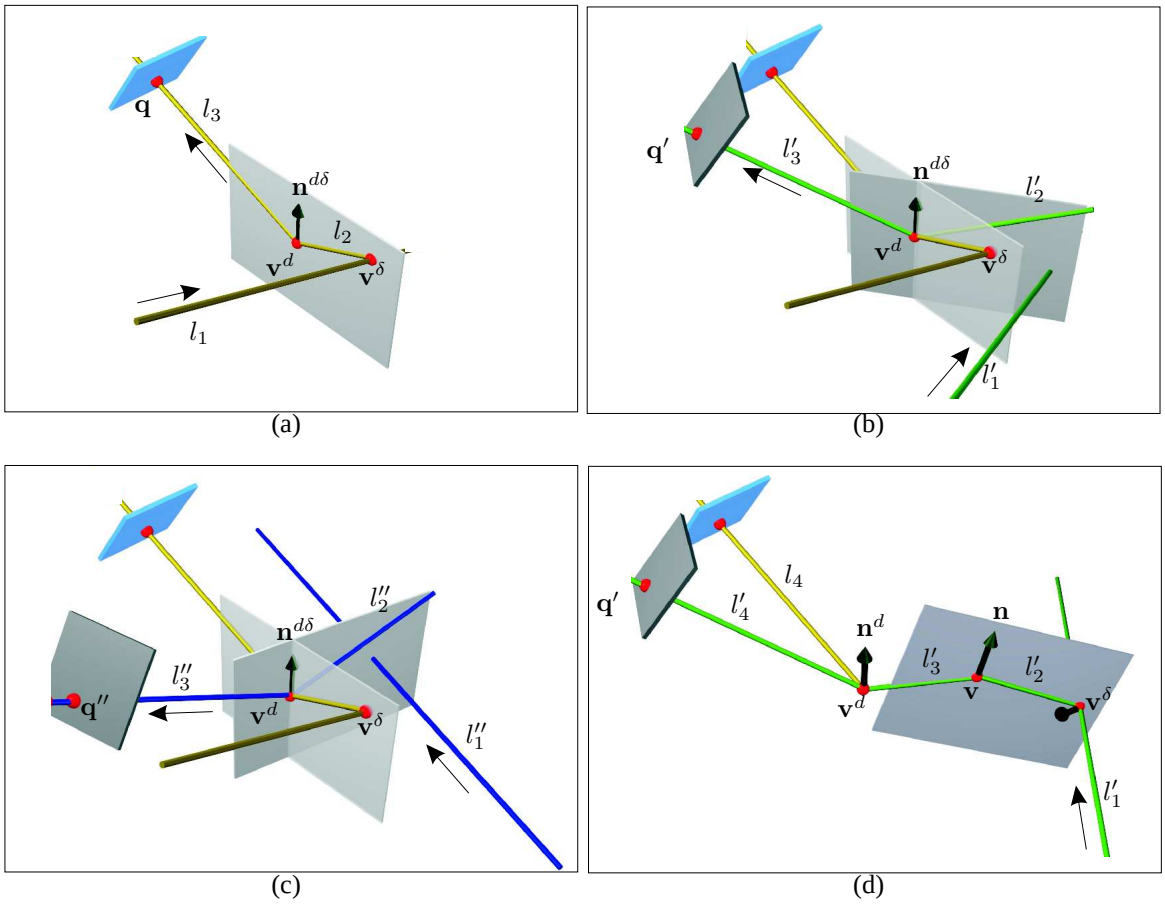


Figure 7: (a)-(c) Path geometries in proof of Proposition 1. (a) Light path of an image point $\mathbf{q}$ in the first viewpoint. The arrow indicates the direction of incoming light. Rays l_1 and l_3 are known but l_2 is not. The shaded plane is the plane of rays l_2 and l_3 and always contains the surface normal, $\mathbf{n}^{d\delta}$. Generically, this plane will not contain ray l_1. (b) Light path of $\mathbf{q}'$ in the second viewpoint, for a given value of d and δ. The path in (a) is also shown. Rays l'_1 and l'_3 are known. Ray l'_2 is uniquely determined by l'_3 and $\mathbf{n}^{d\delta}$. For arbitrary d and δ, the rays l'_1 and l'_2 will not intersect. The dark-shaded plane is the plane of l'_2 and l'_3. (c) Light path of $\mathbf{q}''$ in the third viewpoint. (d) Path geometries in proof for Proposition 2.

To prove the proposition we use two facts. First, since $M = 2$, we know two rays on the light path of every image point. Second, for scenes bounded by a generic (i.e., non-degenerate) surface [40], the light path of almost every pixel, in a measure-theoretic sense, will be non-planar, i.e., the first and last ray of a light path will not lie on the same plane, and therefore these rays will not intersect (Figure 7a). This is because the planarity of a light path is not a stable [37] property—almost any infinitesimal surface deformation, change in viewpoint, or change in the

position of pixel $\mathbf{q}$ will invalidate it.

Now let $\mathbf{q}$ be an arbitrary image point, let l_1, l_2, l_3 be the first, middle, and last ray along its light path, respectively, and let d be a hypothetical depth value assigned to $\mathbf{q}$. We show that in general only isolated d-values can define a consistent light network.

Since l_1 is the first ray on the light path of $\mathbf{q}$, it contains the first vertex of $\mathbf{q}$'s path. Moreover, since this ray is known, there is a one-degree-of-freedom ambiguity in the position of this vertex. We can therefore parameterize its position with a parameter $\delta \in (-\infty, \infty)$. For a given d, each δ-value defines a unique position, $\mathbf{v}^\delta$, for the path's first vertex and, consequently, a unique light path for $\mathbf{q}$. In that path, light initially propagates along l_1, is refracted at $\mathbf{v}^\delta$ and then at $\mathbf{v}^d$, and finally reaches $\mathbf{q}$. From the Deflection Property, only one normal at $\mathbf{v}^d$ can redirect light according to that path for any given value of the refractive index. Now suppose that we fix the refractive index to a specific, but possibly incorrect, value. In this case, it is possible to map every pair (d, δ) to a normal, $\mathbf{n}^{d\delta}$. Moreover, since l_1 and l_3 do not intersect in general, this mapping is a diffeomorphism for almost every $\mathbf{q}$. Note that we can compute $\mathbf{n}^{d\delta}$ for any d and δ because we know l_1 and l_3.

Now let $\mathbf{q}'$ be the perspective projection of point $\mathbf{v}^d$ in the second viewpoint, and let l'_1 and l'_3 be the first and last ray on its light path, respectively (Figure 7b). Rays l'_1 and l'_3 will also not intersect in general. Given a normal $\mathbf{n}^{d\delta}$ and ray l'_3, the Deflection Property tells us that there is a unique ray, l'_2, that (1) passes through $\mathbf{v}^d$ and (2) causes light propagating along l'_2 to be refracted toward $\mathbf{q}'$. This ray is completely determined by $\mathbf{v}^d$, $\mathbf{n}^{d\delta}$, the second viewpoint, and the image point $\mathbf{q}'$. In particular, *there is no geometric constraint between rays l'_1 and l'_2*. It follows that these rays will be in general position, i.e., they will not intersect for an arbitrary choice of d and δ and will not form a light path. Hence, such a choice does not produce a light network for $\mathbf{q}$.

For a given d, there is only an isolated set of δ-values that cause rays l'_1 and l'_2 to intersect. To see this, note that as δ varies over the interval $(-\infty, \infty)$, ray l'_2 traces a ruled surface whose shape has no relation to ray l'_1. Since in general a ray and a surface will only have isolated intersection points [37], and since l'_1 and l'_2 intersect precisely at those points, it follows that for every d there is only a discrete set, Δ^d, of δ-values that produce a light path through $\mathbf{q}'$.

Finally, consider the projection, $\mathbf{q}''$, of $\mathbf{v}^d$ in the third viewpoint (Figure 7c). For a given d, the normals that produce light paths for the first two viewpoints are given by the set $\{\mathbf{n}^{d\delta} \mid \delta \in \Delta^d\}$. For every normal in this set there is a unique ray, l''_2, that passes through point $\mathbf{v}^d$ and forces light propagating along l''_2 to be refracted toward pixel $\mathbf{q}''$. Since the set of normals is discrete, these rays form a discrete family. Moreover, since this family of rays has no relation to ray l''_1 and since

rays in general position have no common intersections, it follows that rays l_1'' and l_2'' will either never intersect, or will intersect for an isolated set of d-values. When the "true" refractive index is known, however, there is at least one such value that produces a consistent light network—the "true" depth. *QED*

Proof of Proposition 1(b): Our proof is a continuation of the proof for part (a). We show that, generically, rays l_1'' and l_2'' will intersect for an isolated set of d-values even when we use an incorrect refractive index. This implies the existence of a depth-refractive index ambiguity: for almost any hypothesized value for the refractive index, there is a depth hypothesis that gives rise to a consistent light network.

More specifically, suppose we fix the refractive index to an arbitrary value ρ and consider the arrangement of rays l_1'' and l_2'' in Figure 7c. As the hypothetical depth value d varies in the range $(-\infty, \infty)$, the vertex $\mathbf{v}^d$, its projection $\mathbf{q}''$ in the third viewpoint, and rays l_1'' and l_2'' will also vary as functions of d. To make this explicit, we parameterize l_1'' and l_2'' by the refractive index and the depth hypothesis that give rise to them, i.e., $l_1''(\rho, d)$ and $l_2''(\rho, d)$ are the rays corresponding to depth hypothesis d for refractive index ρ. For a given ρ, these rays trace a pair of ruled surfaces in 3D. When ρ varies as well, these ruled surfaces span a 3D volume.

We now define the following two sets:

$$\mathcal{M}_1 = \{(x, y, z, \rho, d) \mid (x, y, z) \textit{ is a point on ray } l_1''(\rho, d) \}\ , \tag{2}$$

$$\mathcal{M}_2 = \{(x, y, z, \rho, d) \mid (x, y, z) \textit{ is a point on ray } l_2''(\rho, d) \}\ . \tag{3}$$

Observe that $\mathcal{M}_1$ and $\mathcal{M}_2$ have a non-empty intersection if and only if there is a refractive index ρ such that rays $l_1''(\rho, d)$ and $l_2''(\rho, d)$ intersect for some depth hypothesis d. These two sets are 3-dimensional manifolds in $\mathbb{R}^5$ and, generically, two such manifolds intersect along a 1-dimensional manifold (i.e., a codimension-4 submanifold of $\mathbb{R}^5$ [37]). This manifold represents the depth-refractive index ambiguity. Hence, when the refractive index is unknown, the solution space of the $\langle 3, 2, 2\rangle$-triangulation problem is a 1-dimensional, rather than a 0-dimensional, manifold. *QED*

3.3 The Limits of Light-Path Triangulation

We now prove that light-path triangulation cannot reconstruct general scenes that redirect light more than twice.

Proposition 2 $\langle N, 3, 2\rangle$*-triangulation is intractable.*

Proof: It suffices to prove the proposition for the case where the scene is refractive with a known refractive index and is viewed from $N > 1$ viewpoints. Let d be a hypothetical depth value at $\mathbf{q}$, and let $\mathbf{n}^d$ be an arbitrarily-chosen normal for vertex $\mathbf{v}^d$ (Figure 7d). Given the projection $\mathbf{q}'$ of $\mathbf{v}^d$ in the i-th viewpoint, we will assign coordinates and normals to all remaining vertices on its light path in a way that is consistent with the laws of refraction.

We use the same terminology as in the proof of Proposition 1. For a given d and $\mathbf{n}^d$, there is only one ray, l'_3, that can refract light toward image point $\mathbf{q}'$ (Figure 7d). The second vertex, $\mathbf{v}$, on $\mathbf{q}'$'s light path will lie on that ray. Choose an arbitrary location on the ray for that vertex. To fully define a light path for $\mathbf{q}$, we now need to specify its first vertex. This vertex must lie on the known ray l'_1. As in the proof of Proposition 1, the 3D position, $\mathbf{v}^\delta$, of this vertex can be parameterized by a single parameter δ. Choose an arbitrary value of δ to fix the location of that vertex as well. Now, the Deflection Property tells us that there is a unique normal that will redirect light from l'_2 toward l'_3 at $\mathbf{v}$. Similarly, there is a unique normal that will redirect light from l'_1 toward l'_2 at $\mathbf{v}^\delta$. Hence, we have found an assignment of 3D coordinates and normals for all path vertices that produces a light path for $\mathbf{q}'$. Since we were able to do this for an arbitrary value of the depth d, the triangulation problem's solution space is dense in $\mathbb{R}$. *QED*

3.4 The Power of Global Shape Recovery

The fact that light-path triangulation is intractable for scenes with long light paths does not necessarily mean that reconstruction of such scenes is hopeless. Intuitively, light-path triangulation operates at a completely local level—for any two points on the same image plane, it attempts to reconstruct the associated light networks independently of each other. So what if we had a procedure that reasoned about multiple light networks simultaneously? Here we briefly sketch a partial answer to this question: we show that a sufficiently large collection of viewpoints does contain enough information to reduce shape ambiguities to a discrete set. Although this existence result does not point to any algorithms, it does suggest that, with enough images, we can test with reasonable confidence the validity of a hypothesized 3D scene model:

Proposition 3 *Given an arrangement of viewpoints for which there is a constant K such that (1) every scene point is intersected by at least $3(K-1)$ light paths of length $\leq K$ and (2) the first and last ray of all these paths is known, the location of each scene point is constrained to a 0-dimensional solution manifold.*

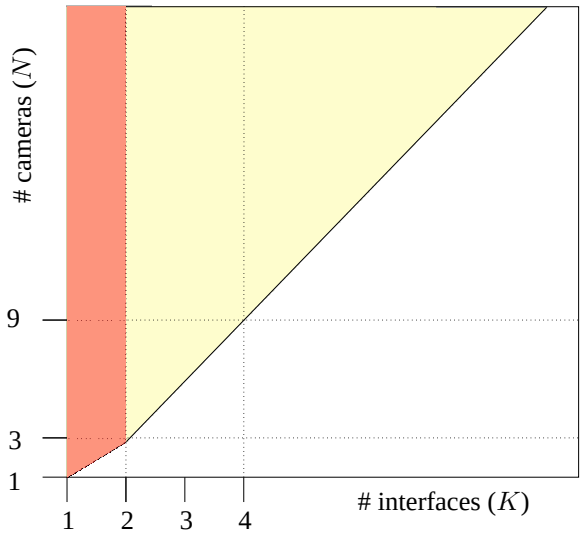
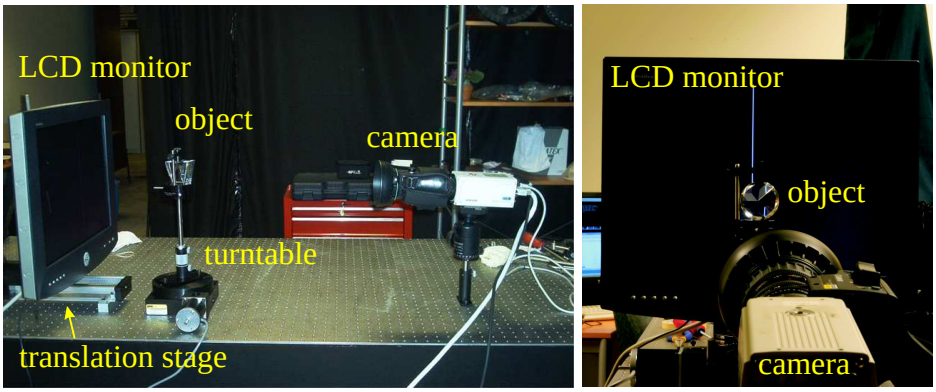


Figure 8: The space of solvable specular scene reconstruction problems. For values of N, K in the dark-colored region (red), reconstruction is possible by $\langle N, K, 2\rangle$-triangulation, according to Theorem 1. When N, K are in the light-colored region (yellow), $\langle N, K, 2\rangle$-triangulation is intractable, but Proposition 3 tells us that reconstruction may still be possible using a global approach.

Intuitively, Proposition 3 gives us a lower bound on the number of viewpoints we need for shape verification: for light paths of maximum length K, each scene point must project indirectly to least $3(K-1)$ viewpoints. We prove this result inductively, using Proposition 1 both as the base case and for proving the inductive step.

Proof sketch: The base case is covered by Proposition 1. To show that it holds for $K = k$, we define a partitioning of the scene into k "layers," $\mathcal{S}_1, \ldots, \mathcal{S}_k$, where the i-th layer contains scene points that (1) participate as the i-th vertex on at least one light path with known first and last rays and length $\leq k$, and (2) they never participate as a lower-numbered vertex on such a path. The proof is restricted to scenes where each of the $\mathcal{S}_i$ is a smooth manifold. Now assume that we know the 3D position and surface normal of all points in $\mathcal{S}_k$. This assumption uniquely determines the second-to-last ray of all k-vertex light paths with known first and last ray. We now apply the proposition with $K = k-1$ to the scene defined by the first $k-1$ layers, using $3(k-2)$ of the $3(k-1)$ light paths that cross each point and have known first, last and second-to-last ray. It follows that if $\mathcal{S}_k$ is known, each point on the remaining layers is constrained to a 0-dimensional solution manifold. We can now use a proof identical to that of Proposition 1 to show that the remaining three light paths that cross each point and were not used in the inductive step, constrain the 3D position of each point on layer $\mathcal{S}_k$ to a 0-dimensional manifold as well. *QED*



(a) (b)

Figure 9: Acquisition setup for $\langle 3,2,2\rangle$-triangulation. (a) A linear translation stage moves the LCD monitor in a forward/backward direction. To change viewpoint, the object is rotated by a computer-controlled rotation stage. (b) During image acquisition, the LCD displays a black background with a moving horizontal or vertical stripe. This stripe is used to establish a correspondence between pixels in the image and the 3D locations on the LCD monitor that project to those pixels indirectly.

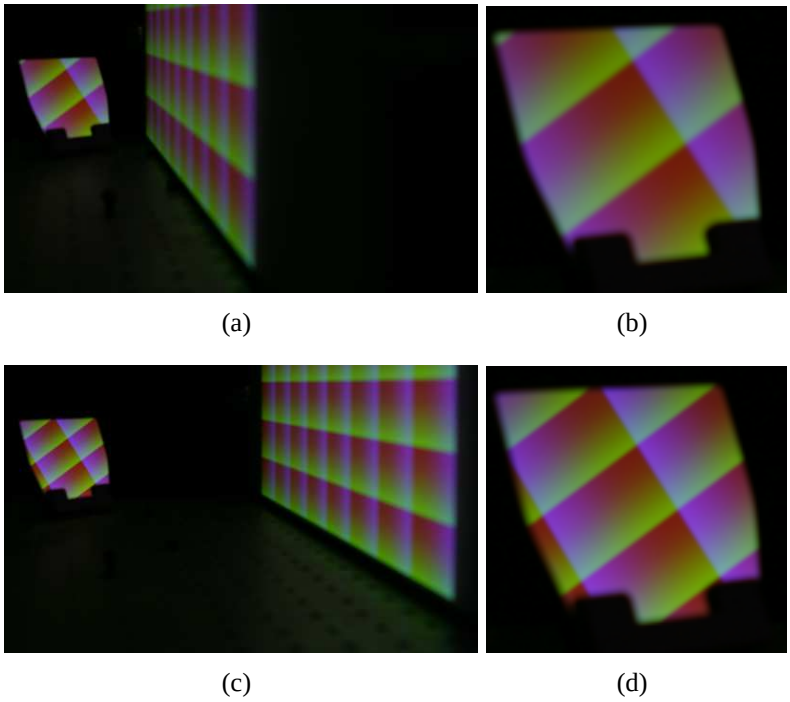
4 Experimental Results

While our emphasis in this article is on the underlying theory, we performed initial experiments on all three tractable instances of light-path triangulation. Below we briefly present results for two of those instances, namely $\langle 1,1,2\rangle$-triangulation and $\langle 3,2,2\rangle$-triangulation. For detailed case studies of $\langle 2,1,1\rangle$-triangulation and $\langle 3,2,2\rangle$-triangulation, including algorithms, implementation details and more results, see [31, 41] and [42], respectively.

For the experiments below, we used a 720×484-pixel Sony DXC-9000 video camera for image acquisition and a DELL 1600×1200 LCD display for displaying reference patterns, whose position was under computer control (Figures 9 and 10). To calibrate the camera with respect to the plane of the LCD display, we used the Matlab Calibration Toolbox [32], and used an environment matting procedure [5] to find the correspondence between image pixels and pixels on the display. The display was then translated by a known amount and the procedure was repeated, giving us two known 3D reference points per image pixel (Figure 10).

4.1 Reconstructing mirrors by $\langle 1,1,2\rangle$-triangulation

We used the arrangement in Figure 5 (left) and Figure 10. A key feature of $\langle 1,1,2\rangle$-triangulation is that reconstruction accuracy largely depends on the accuracy of camera calibration, not on the



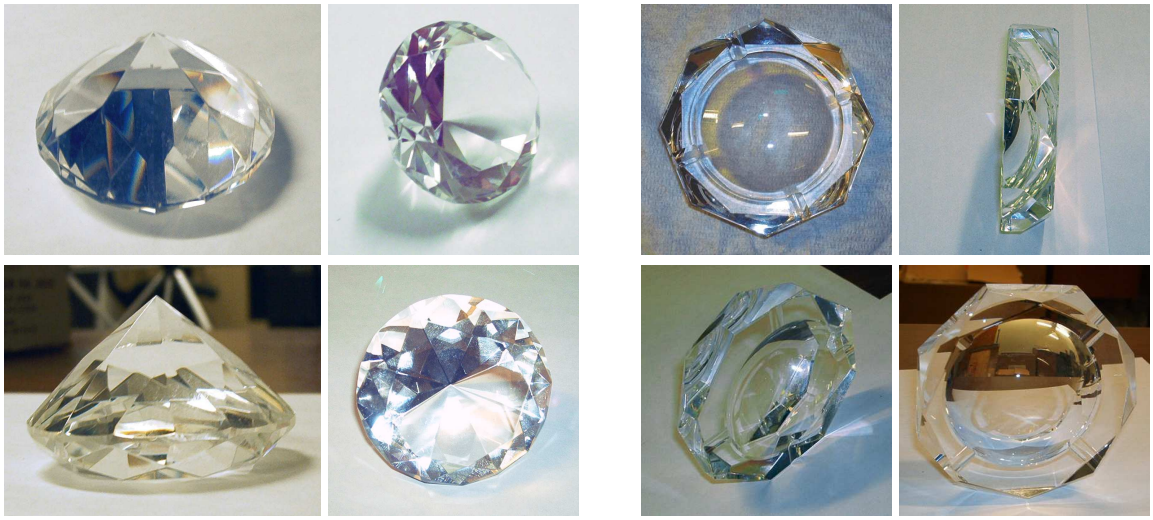
(a) (b)

(c) (d)

Figure 10: Acquisition setup for $\langle 1, 1, 2 \rangle$-triangulation. (a) The mirror used in our experiment is on the far left of the image, reflecting the pattern on the LCD monitor, visible on the right. (b) Close-up of the mirror. (c) The LCD monitor is moved to a second position, farther from the mirror, without changing the camera's viewpoint. (d) Close-up of the mirror for the new monitor position. Note that the reflection on the mirror changed between (b) and (d) because the light path through a given camera pixel passes through a different reference point on the monitor. The pattern on the monitor is for illustration only; in practice, camera-to-LCD correspondences are established by displaying a moving stripe, as in Figure 9b.

shape of the object being reconstructed. We therefore concentrated on evaluating the accuracy of the depths and normals computed individually for each pixel, with an object whose ground-truth shape was known very accurately: a 130×230mm front-surface mirror with $\frac{1}{4}$-wavelength flatness. To determine the mirror's plane, we digitized several points on it with a FaroArm Gold touch probe, whose single-point measurement accuracy is $\pm$0.05mm, and then fit a plane through these points. The mirror was placed about 1.5m away from the camera.

To compute the depth d at a pixel, we simply intersected the first and last ray along its light path (see Eq. (1) and Figure 5(left)). The bisector of these rays gave us the surface normal. This computation was done at each of 301,082 pixels in the image, giving us an equal number of 3D position and normal measurements. No smoothing or post-processing was applied. The RMS



(a) (b)

Figure 11: Views of two objects used in our $\langle 3, 2, 2\rangle$-triangulation experiments. (a) A diamond-shaped glass object. (b) A glass ashtray.

distance of the reconstructed 3D points from the ground-truth plane was 0.644mm, equivalent to a single-point accuracy of roughly 99.96% of the camera-to-object distance. To assess the accuracy of reconstructed normals, we measured the angle between each computed normal and the ground-truth normal; the mean error was 0.182 degrees, showing that single-point orientation measurements were also highly accurate. We emphasize that these accuracies were obtained without using any information about the scene's shape and without combining measurements from multiple pixels.

4.2 Reconstructing glass objects by $\langle 3, 2, 2\rangle$-triangulation

We used the arrangement in Figure 6 and Figure 9. Since this triangulation requires three or more viewpoints, we place objects on a turntable between the LCD and the camera and compute the correspondence between image pixels and pixels on the monitor for each object rotation.

Figure 11 shows two of the objects used in our experiments, a diamond-shaped glass ornament and a glass ashtray. We used the same viewing configuration for both objects: a total of seven viewpoints, corresponding to $\pm 30, \pm 20, \pm 10$ and 0-degree rotations (Figure 12). Both objects extended 3 to $6cm$ in depth, roughly $1.2m$ away from the camera. To reconstruct them, we used all available views and solved a $\langle 7, 2, 2\rangle$-triangulation problem independently for every pixel in the 0-degree viewpoint. For each such pixel, our implementation performed a search in (d, δ)-space

for a pair of values that produce a consistent light network, i.e., a network whose light paths are consistent with the laws of refraction for all viewpoints (Section 3.2 and Figures 7a-c). These values were then refined in a non-linear optimization stage. See [42] for details. Since the light network of a pixel contains eight vertices, the algorithm reconstructs eight points and eight normals per pixel—one on the object's front surface and seven more on the back (Figure 6). Importantly, since we used more viewpoints than the minimum three required, the reconstruction was over-constrained and allowed estimation of the objects' refractive index as well.

4.2.1 "Diamond" scene

This object, shown in Figure 11a, has many surface features that make reconstruction especially challenging. These include numerous planar facets on both the front and the back surfaces, that produce complex light paths; many surface discontinuities; non-planar front and back surfaces; and a sharp, protruding tip where surface orientation is degenerate.

Figures 13 and 14 show reconstruction results. The object's index of refraction was estimated to be 1.55. The maps for the normals' slant and tilt angles suggest that the object's surface orientation was highly consistent across different pixels within a facet, even though light paths for different pixels were reconstructed completely independently, and no smoothing or post-processing was applied. Moreover, since light path triangulation computations are performed independently for each pixel in the reference view, surface normals were reconstructed accurately even for pixels near the diamond's tip, where the surface orientation field is singular. Also observe that, as a side-effect, we obtain an automatic segmentation of the scene into smooth segments. This is because image-to-LCD correspondences cannot be established at the precise location of a normal discontinuity and, hence, those pixels were not reconstructed.

To further assess the precision of our reconstruction, we measured the consistency of normals and depths within each planar facet. These quantitative measurements are shown in Figures 15a and 15c. They show that individually-reconstructed normals within a facet are consistent to within a few degrees, while depth measurements, which seem to produce a noisier map, show deviations of about 0.1% of the object-to-camera distance. These results confirm our basic theory and suggest that it is possible to recover detailed shape information for refractive objects without any knowledge of their shape, despite the complexity of image formation.

4.2.2 "Ashtray" scene

Our second scene, the ashtray shown in Figure 11b, has very different surface properties from the diamond. Its front surface is composed of a large, smooth and concave region in the center; a planar region in the periphery; and several small planar facets adjacent to it. Its back surface, on the other hand, is a large planar base. Because of the prominent concavity, this scene would be impossible to reconstruct accurately with a silhouette-based method (e.g., [4]).

The ashtray's index of refraction was estimated to be 1.54. Reconstruction results are shown in Figures 16 and 17. As in the previous example, depth and normal measurements are computed independently for each pixel in the reference view, and no smoothing or post-processing was applied. To assess these results further, we examine the precision of normal estimates for the four surface regions marked A, B, C and D in Figure 15b.

In region A, both the normal and the depth estimates change very smoothly, producing a highly-consistent concave surface. This is confirmed quantitatively in the table of Figure 15d, where normal variation within small pixel neighborhoods was measured to be about two degrees. This suggests that reconstruction quality for this non-planar region is on par with that of the "diamond" scene.

The reconstruction results for the remaining regions show significant variations in either the depth or the normal estimates of individual pixels. These regions correspond to *critical configurations* of $\langle 3, 2, 2 \rangle$-triangulation and illustrate some of the limitations of our current implementation and of our experimental setup. More specifically, the light path of pixels in region D intersects the object along two parallel planes, one on the front and one on the back side of the ashtray. In this case, the entire light path is planar, violating the generic light path non-planarity assumption that is exploited by $\langle 3, 2, 2 \rangle$-triangulation (Section 3.2 and Figure 7a). It is easy to show that in the presence of this "parallel-plane degeneracy" we can compute the planes' normal and their inter-plane distance but we cannot compute absolute depth [42]. This fact is confirmed by our results: the table of Figure 15d and the normal maps in Figure 16 show that normals in region D are highly consistent. In contrast, the reconstructed depth map is not accurate in region D (e.g., see Figure 16a and Figure 17, bottom right).

The pixels in regions B and C, on the other hand, correspond to critical configurations caused by camera placement. In region B, the camera's motion and the normal of the front and back surfaces are all approximately coplanar; in region C, all surface points are approximately on the object's axis of rotation. Both cases result in a reduced set of geometric constraints that allow either the normal slant angle or the normal tilt angle to be determined uniquely, but not both [42].

This fact is consistent with the normal maps in Figure 16, as well as with the normal precision measurements shown in Figure 15d.

While our current implementation did not seek to detect or overcome the above critical configurations, the process for doing so is fairly straightforward. Pixels whose light path is planar can be easily detected because both the first and the last ray on their light path is known. It is therefore possible to determine *a priori* whether or not a pixel has a planar light path by checking whether or not these two rays intersect in 3D. In addition, degeneracies caused by camera placement can be avoided by relying on a 2D set of input viewpoints for reconstruction, rather than single-axis rotations. We should note, however, that the above degeneracies may not be exhaustive—the theoretical problem of characterizing all possible critical configurations of $\langle 3, 2, 2 \rangle$-triangulation is largely open, and is beyond the scope of this initial study.

5 Concluding Remarks

This article introduced *light-path triangulation* as a computational framework for analyzing the reconstruction of general specular scenes from photographs. We have shown that this framework provides a unified analysis of transparent and mirror-like scenes, leading to "computability" results and practical algorithms that apply to both problems. On the theoretical side, our main contribution was to characterize the set of reconstructible scenes in a way that depends only on the number of vertices along a light path. This led to three key results. First, we derived a simple algorithm for computing the depth map of a specular scene from a single viewpoint, when the scene redirects incoming light just once. Second, for unknown scenes where incoming light undergoes *two* refractions or reflections, we showed that three viewpoints are sufficient to enable reconstruction in the general case. To our knowledge, no other algorithms exist for reconstructing light paths of this complexity. Third, we showed that it is impossible to reconstruct individual light paths when light is redirected more than twice. This effectively established a "computability" limit, bounding the complexity of scenes that can be reconstructed by an efficient, stereo-like algorithm.

While our preliminary experimental results are promising, many practical questions remain open. These include (1) how to best compute correspondences between reference points and pixels, (2) how to reconcile point and normal measurements, (3) how to find the optimal depth at a pixel, (4) how to identify all critical configurations for specular reconstruction, and (5) how to develop algorithms and acquisition procedures that either avoid these configurations or operate robustly in their presence. Finally, our theoretical analysis can be thought of as a "worst-case" scenario for

reconstruction, where no constraints are placed on nearby scene points. Since real scenes exhibit spatial coherence, it might be possible to incorporate this constraint into an algorithm that remains tractable even for scenes that refract light more than twice.

6 Acknowledgements

This work was supported in part by the National Science Foundation under Grant No. IRI-9875628, by the Natural Sciences and Engineering Research Council of Canada under the RG-PIN and PGS-M programs, by a fellowship from the Alfred P. Sloan Foundation, by an Ontario Premier's Research Excellence Award and by Microsoft Research. The authors would also like to thank Chris Trendall for his implementation of $\langle 2, 1, 1 \rangle$-triangulation, as well as Kurt Akeley, Steven Lin, Allan Jepson, Aaron Hertzmann and the anonymous reviewers for their many helpful comments on versions of this manuscript.

References

[1] O. Faugeras and R. Keriven, "Complete dense stereovision using level set methods," in *Proc. 5th European Conf. on Computer Vision*, pp. 379–393, 1998.

[2] A. Hertzman and S. M. Seitz, "Shape and materials by example: A photometric stereo approach," in *Proc. Computer Vision and Pattern Recognition Conf.*, pp. 533–540, 2003.

[3] T. Zickler, P. N. Belhumeur, and D. J. Kriegman, "Helmholz stereopsis: Expoiting reciprocity for surface reconstruction," in *Proc. 7th. European Conf. on Computer Vision*, pp. 869–884, 2002.

[4] W. Matusik, H. Pfister, R. Ziegler, A. Ngan, and L. McMillan, "Acquisition and rendering of transparent and refractive objects," in *Proc. 13th Eurographics Workshop on Rendering*, pp. 267–278.

[5] D. E. Zongker, D. M. Werner, B. Curless, and D. H. Salesin, "Environment matting and compositing," in *Proc. ACM SIGGRAPH*, pp. 205–214, 1999.

[6] M. Baba, K. Ohtani, and M. Imai, "New laser rangefinder for three-dimensional shape measurement of specular objects," *Optical Engineering*, vol. 40, no. 1, pp. 53–60, 2001.

[7] M. A. Halstead, B. A. Barsky, S. A. Klein, and R. B. Mandell, "Reconstructing curved surfaces from specular reflection patterns using spline surface fitting of normals," in *Proc. ACM SIGGRAPH*, pp. 335–342, 1996.

[8] J. Y. Zheng and A. Murata, "Acquiring a complete 3d model from specular motion under the illumination of circular-shaped light sources," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 8, pp. 913–920, 2000.

[9] M. Tarini, H. P. A. Lensch, M. Goesele, and H.-P. Seidel, "3D acquisition of mirroring objects using striped patterns," *Graphical Models*, vol. 67, no. 4, pp. 233-259, 2005.

[10] X. Zhang and C. S. Cox, "Measuring the two-dimensional structure of a wavy water surface optically: A surface gradient detector," *Experiments in Fluids*, vol. 17, pp. 225–237, 1994.

[11] S. Agarwal, S. P. Mallick, D. Kriegman, and S. Belongie, "On refractive optical flow," in *Proc. 8th European Conf. on Computer Vision*, pp. 483–494, 2004.

[12] J. Höhle, "Reconstruction of the underwater object," *Photogrammetric Engineering*, pp. 948–954, 1971.

[13] H.-G. Maas, "New developments in multimedia photogrammetry," in *Optical 3D Measurement Techniques III*, Wichmann Verlag, 1995.

[14] M. Ben-Ezra and S. Nayar, "What does motion reveal about transparency?," in *Proc. 9th Int. Conf. on Computer Vision*, pp. 1025–1032, 2003.

[15] S. Savarese and P. Perona, "Local analysis for 3d reconstruction of specular surfaces - part ii," in *Proc. 7th European Conf. on Computer Vision*, pp. 759–774, 2002.

[16] W. C. Keller and B. L. Gotwols, "Two-dimensional optical measurement of wave slope," *Applied Optics*, vol. 22, no. 22, pp. 3476–3478, 1983.

[17] B. Jähne, J. Klinke, and S. Waas, "Imaging of short ocean wind waves: a critical theoretical review," *J. Opt. Soc. Am. A.*, vol. 11, no. 8, pp. 2197–2209, 1994.

[18] H. Murase, "Surface shape reconstruction of an undulating transparent object," in *Proc. 3rd Int. Conf. on Computer Vision*, pp. 313–317, 1990.

[19] K. Ikeuchi, "Determining surface orientations of specular surfaces by using the photometric stereo method," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 3, no. 6, pp. 661–669, 1981.

[20] J. Wang and K. J. Dana, "A novel approach for texture shape recovery," in *Proc. 9th Int. Conf. Computer Vision*, pp. 1374–1380, 2003.

[21] D. Miyazaki, M. Kagesawa, and K. Ikeuchi, "Transparent surface modeling from a pair of polarization images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 26, no. 1, pp. 73–82, 2004.

[22] B. Trifonov, D. Bradley, and W. Heidrich, "Tomographic reconstruction of transparent objects," in *Proc. 17th Eurographics Symp. on Rendering*, pp. 51–60, 2006.

[23] J. Sharpe, U. Ahlgren, P. Perry, B. Hill, A. Ross, J. Hecksher-Sorensen, R. Baldock, and D. Davidson, "Optical projection tomography as a tool for 3d microscopy and gene expression studies," *Science*, vol. 296, pp. 541–545, 2002.

[24] S. Hasinoff and K. N. Kutulakos, "Photo-consistent 3d fire by flame sheet decomposition," in *Proc. 9th Int. Conf. on Computer Vision*, pp. 1184–1191, 2003.

[25] M. D. Grossberg and S. K. Nayar, "A general imaging model and a method for finding its parameters," in *Proc. 8th Int. Conf. Computer Vision*, vol. 2, pp. 108–115, 2001.

[26] R. Pless, "Two view discrete and differential constraints for generalized imaging systems," in *Proc. of the IEEE Workshop on Omnidirectional Vision*, pp. 53–59, 2002.

[27] A. C. Sanderson, L. E. Weiss, and S. K. Nayar, "Structured highlight inspection of specular surfaces," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, no. 1, pp. 44–55, 1988.

[28] T. Bonfort and P. Sturm, "Voxel carving for specular surfaces," in *Proc. 9th Int. Conf. on Computer Vision*, pp. 591–596, 2003.

[29] A. Blake, "Specular stereo," in *Int. Joint Conf. on Artificial Intelligence*, pp. 973–976, 1985.

[30] M. Oren and S. K. Nayar, "A theory of specular surface geometry," in *Proc. 5th Int. Conf. Computer Vision*, pp. 740–747, 1995.

[31] N. Morris and K. N. Kutulakos, "Dynamic refraction stereo," in *Proc. 10th Int. Conf. Computer Vision*, pp. 1573–1580, 2005.

[32] J.-Y. Bouguet, "MATLAB camera calibration toolbox." "http://www.vision.caltech.edu/bouguetj/calib_doc/".

[33] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi, "Efficiently combining positions and normals for precise 3d geometry," in *Proc. ACM SIGGRAPH*, pp. 536–543, 2005.

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

[35] S. K. Nayar, K. Ikeuchi, and T. Kanade, "Surface reflection: Physical and geometrical perspectives," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 7, pp. 611–634, 1991.

[36] J. M. Gluckman and S. K. Nayar, "Planar catadioptric stereo: geometry and calibration," in *Proc. Computer Vision and Pattern Recognition Conf.*, pp. 22–28, 1999.

[37] V. Guillemin and A. Pollack, *Differential Topology*. Prentice-Hall, 1974.

[38] K. N. Kutulakos and E. Steger, "A theory of and refractive and specular 3d shape by light-path triangulation," in *Proc. 10th Int. Conf. on Computer Vision*, pp. 1448–1455, 2005.

[39] T. Bonfort, P. Sturm, and P. Gargallo, "General specular surface triangulation," in *Proc. Asian Conference on Computer Vision*, pp. 872–881, 2006.

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

[41] N. J. W. Morris, "Image-based water surface reconstruction with refractive stereo," Master's thesis, Department of Computer Science, University of Toronto, 2004. http://www.dgp.toronto.edu/ nmorris/Water/thesis.pdf.

[42] E. Steger, "Reconstructing transparent objects by refractive light-path triangulation," Master's thesis, Department of Computer Science, University of Toronto, 2006. http://www.cs.toronto.edu/ esteger/thesis.pdf.

[43] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, pp. 381–395, June 1981.

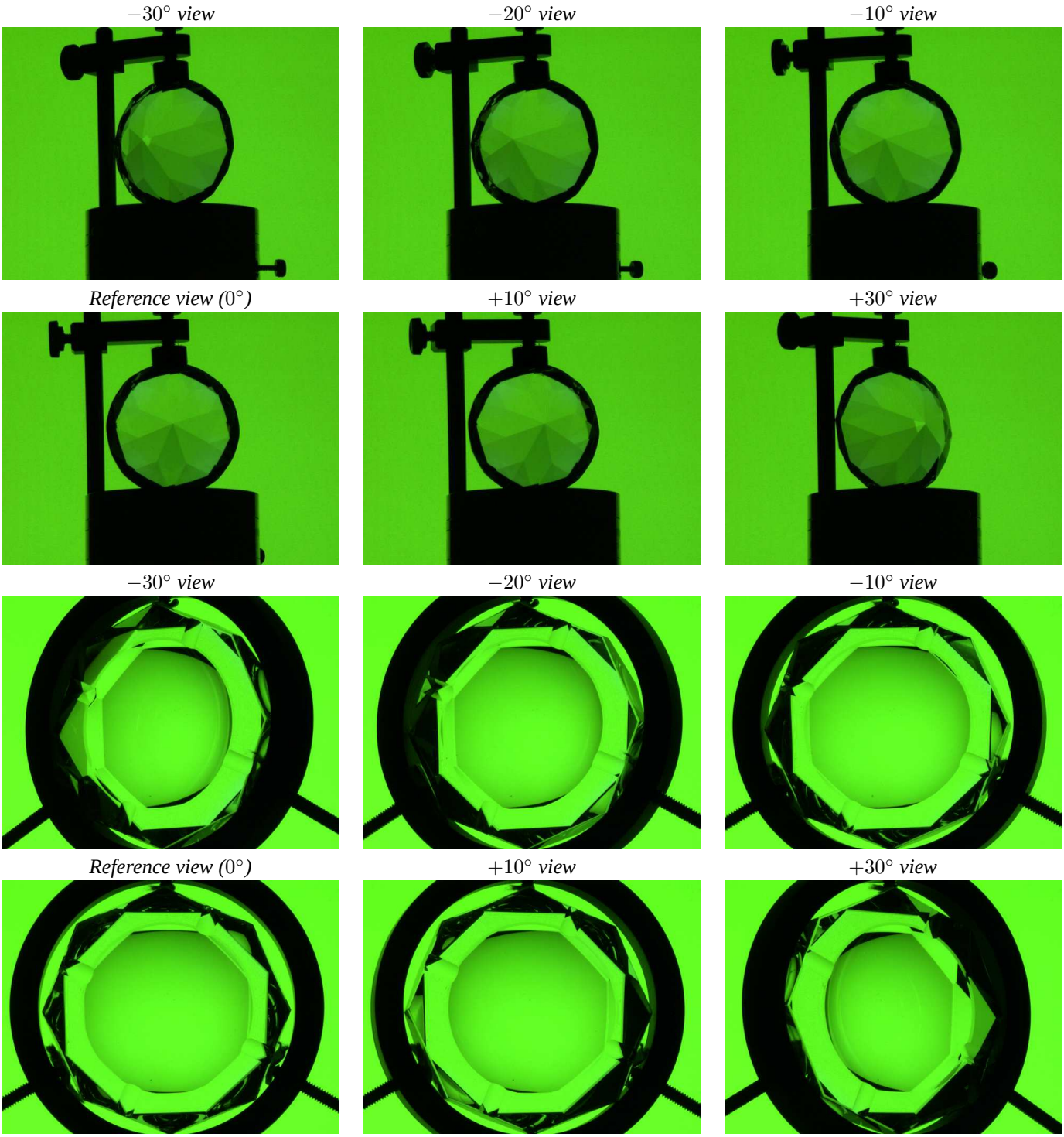


Figure 12: Six of the seven input viewpoints used for reconstruction. During image acquisition, the green background is replaced by a moving, one-pixel-wide horizontal or vertical stripe.

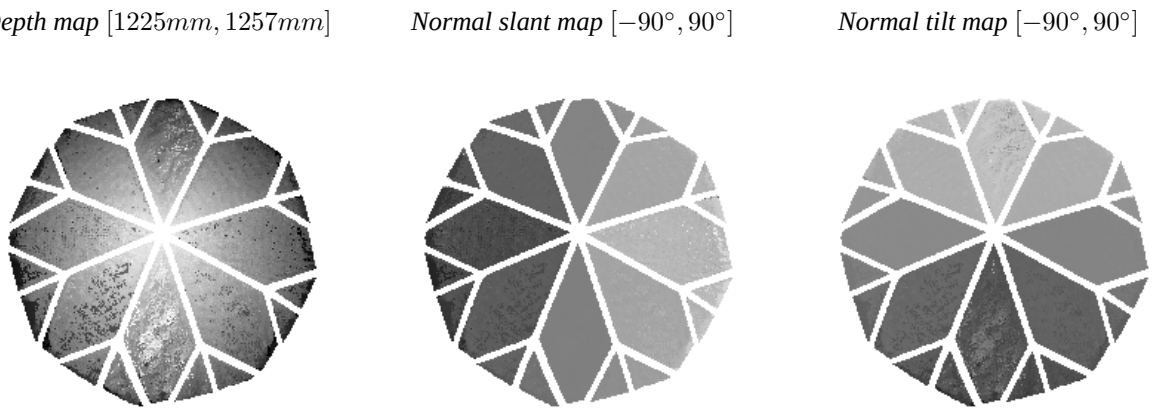


Figure 13: Reconstruction results for the "diamond" scene. (a) Depth map for the reference view. Gray-scale values are mapped to the indicated range (white=near, black=far). (b, c) Reconstructed normal maps for the reference view. Gray-scale values correspond to the slant and tilt angle, respectively, of each reconstructed normal.

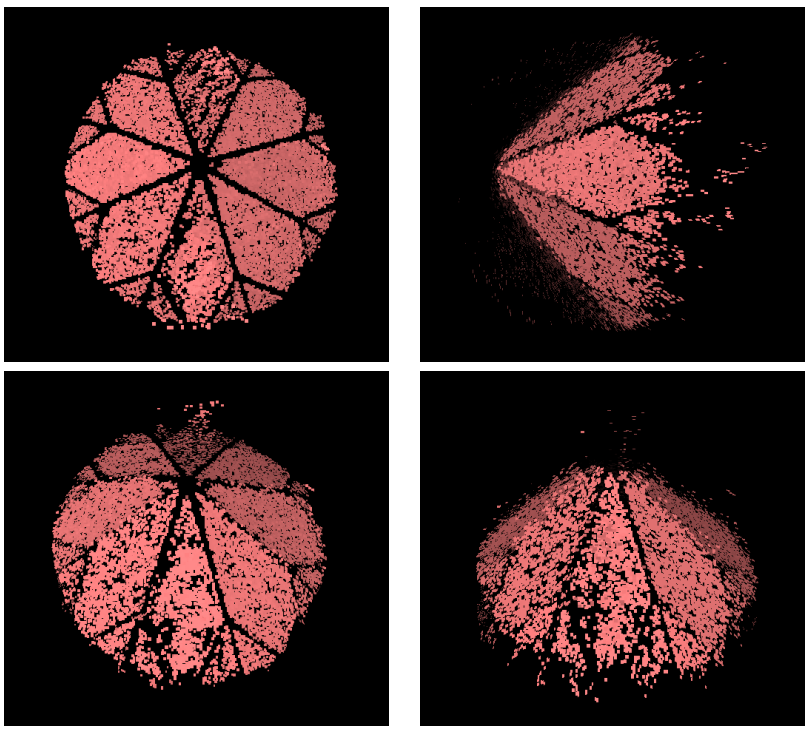
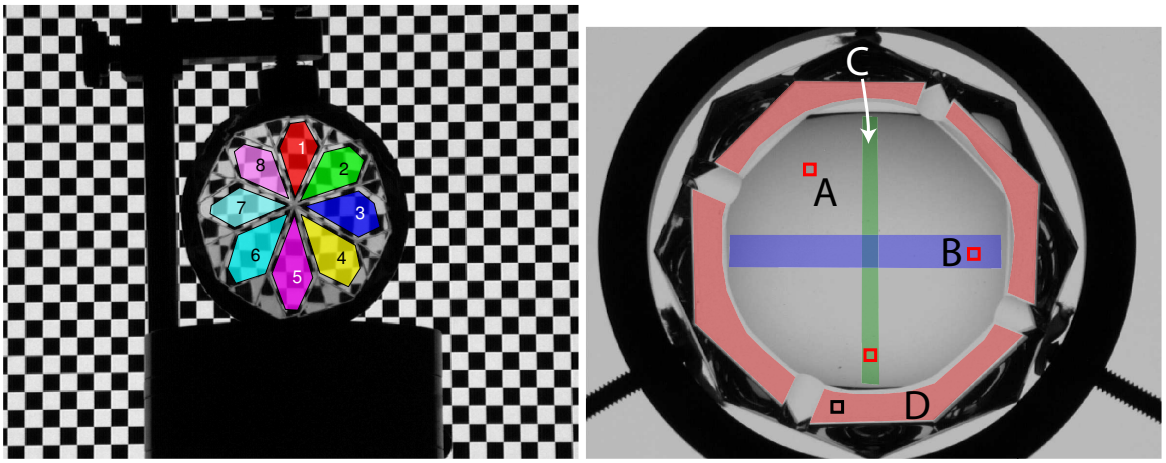


Figure 14: Views of the reconstructed front surface of the "diamond" scene. For each pixel in the reference view, we render a shiny square patch (i.e., a surfel) whose depth is given by the depth map in Figure 13a and whose orientation is given by the normal maps in Figures 13b and 13c.



(a) (b)

Facet label (diamond)	1	2	3	4	5	6	7	8
Mean normal error (degrees)	4.46	1.70	2.80	3.15	8.32	6.42	4.14	1.53
Median normal error (degrees)	2.99	1.21	1.83	1.98	6.34	4.18	2.87	1.11
Mean position error (mm)	2.10	0.74	1.60	1.69	2.18	3.40	1.84	0.86
Median position error (mm)	0.71	0.30	0.45	0.47	1.58	0.65	0.57	0.30
RANSAC position inliers (%)	35.9	64.0	52.8	50.1	18.8	40.3	45.1	67.5

(c)

Region label (ashtray)	A	B	C	D
Mean normal error (degrees)	1.79	10.28	6.64	0.90
Median normal error (degrees)	1.63	7.51	6.75	0.21
Mean slant error (degrees)	1.99	17.20	0.82	1.00
Median slant error (degrees)	1.79	12.12	0.70	0.21
Mean tilt error (degrees)	2.35	2.46	10.63	0.92
Median tilt error (degrees)	2.08	2.24	10.77	0.17
Global mean normal error (degrees)				1.13
Global median normal error (degrees)				0.56

(d)

Figure 15: Precision measurements. (a, c) Measurements for facets of the "diamond" scene: (a) Colored polygons indicate the pixels contributing to a facet's measurements. (c) To assess normal variations within a facet, we compute the mean normal across all pixels in a facet and measure the angle between the normal at each pixel and the facet's mean normal. To assess positional variations, we fit a plane to the 3D point measurements using RANSAC [43] with an inlier threshold of $0.5mm$, and then measure the distance of each reconstructed pixel from this plane. (b, d) Measurements for regions of the "ashtray" scene: (b) Decomposing the ashtray's front surface into four major regions. Region A contains all concave points that do not lie in regions B or C. (d) To assess local normal variations within each region, we compute the mean normal in each 3×3 pixel neighborhood and measure the angle between the normal at each pixel in that neighborhood and the neighborhood's mean normal. The table reports aggregate angle measurements across all neighborhoods in a region (i.e., mean or median angle over all neighborhoods). For region D, which is globally planar, we also computed the mean normal across all pixels in the region and measured the angle between the normal at each pixel in D and the region's mean normal.

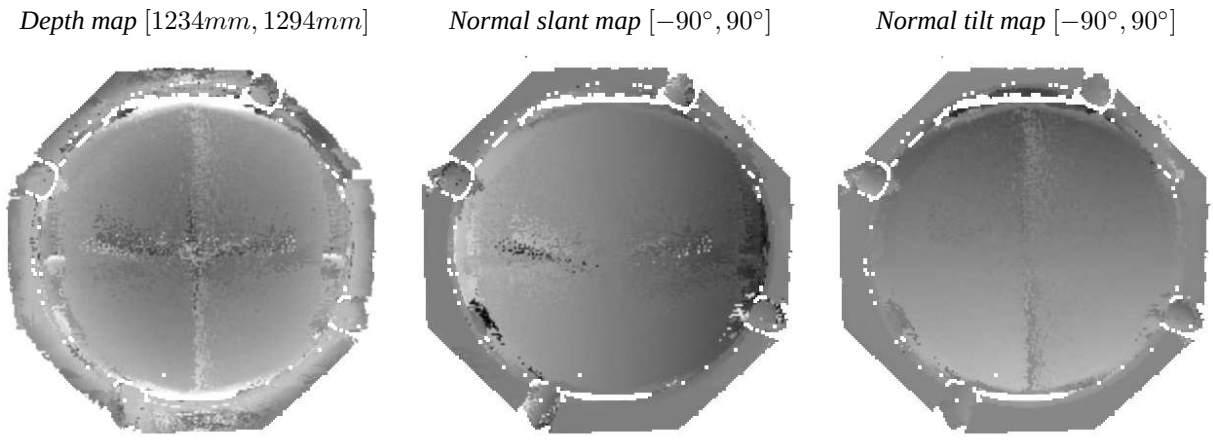


Figure 16: Reconstruction results for the "ashtray" scene. (a) Depth map for the reference view. (b, c) Reconstructed normal maps for the reference view.

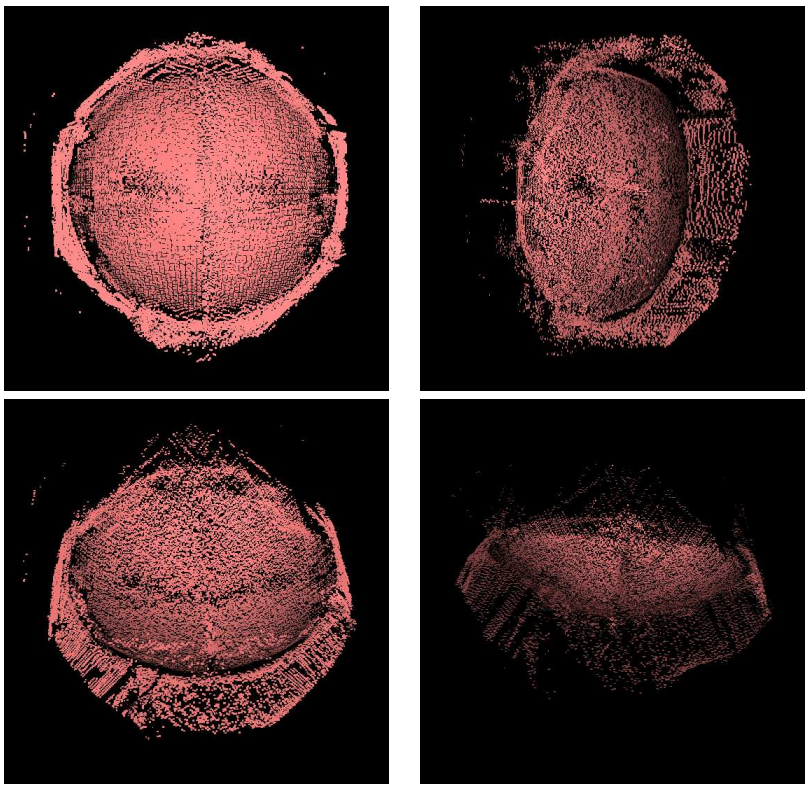


Figure 17: Views of the reconstructed front surface of the "ashtray" scene. For each pixel in the reference view, we render a shiny square patch (i.e., a surfel) whose depth is given by the depth map in Figure 16a and whose orientation is given by the normal maps in Figures 16b and 16c.