



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

Kiriakos N. Kutulakos\*      Eron Steger

University of Toronto

## 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.

## 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 pro-

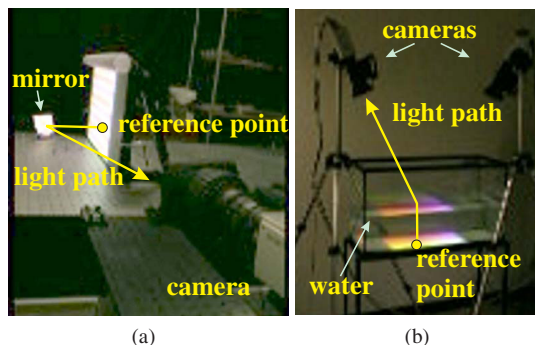


Figure 1. Viewing a known reference point indirectly via (a) an opaque specular scene (a mirror) and (b) a transparent specular scene (a volume of water).

jection 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. This

\*Part of this research was conducted while the author was serving as a Visiting Scholar at Microsoft Research Asia.

has prompted a variety of assumptions, including approximate planarity [17–19], surface smoothness [15], integrability [9], 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, research on reconstructing glass objects has relied on either a silhouette-based approach [4], where an object’s specular properties are not exploited for reconstruction, or on analyzing the polarization of light specularly reflected from their surface [21]. 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.

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, 22, 23] 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* [24–27] 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 [28] 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.

While 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, 29]). 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, 30].

## 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 [31, 32]—*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.<sup>1</sup> 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  $\mathbf{q}$ .

<sup>1</sup>See [33] for a camera-mirror arrangement that forces scene points to indirectly project twice onto the image plane.

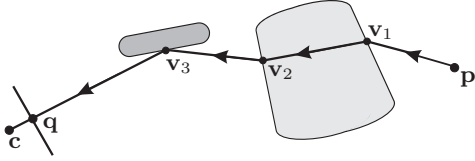


Figure 2. An example light path. The dark gray region denotes a mirror-like object and the light gray region a transparent object. Here, the light path from  $p$  intersects three surfaces before reaching point  $q$  on the image plane, and therefore has three vertices,  $v_1$ ,  $v_2$  and  $v_3$ , and four rays. In light-path triangulation, the coordinates of  $c$ ,  $q$  and  $p$  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.<sup>2</sup> 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  $q$  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,  $v^d$ , along the ray through the selected image point. If that depth is correct,  $v^d$  would redirect light toward all  $N$  viewpoints in a way that is consistent with the laws of reflection and refraction, 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  $v^d$  (Figure 3). These paths define a graph, that we call the *light network* for depth  $d$ . The network connects the  $N$  perspective projections of  $v^d$  to their corresponding reference points.

**Definition 1 (Consistent Light Network)** The light network for depth  $d$  is *consistent* if we can assign a normal to  $v^d$  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 ( $\langle N, K, M \rangle$ -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.

<sup>2</sup>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.

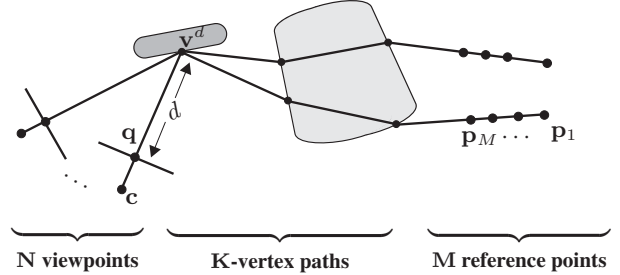


Figure 3. Basic geometry of  $\langle N, K, M \rangle$ -triangulation.

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.<sup>3</sup>
- **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.

<sup>3</sup>Recall that a smooth map,  $f$ , between two manifolds is a local diffeomorphism at a point  $p$  if its derivative,  $df_p$ , is one-to-one and onto [34].

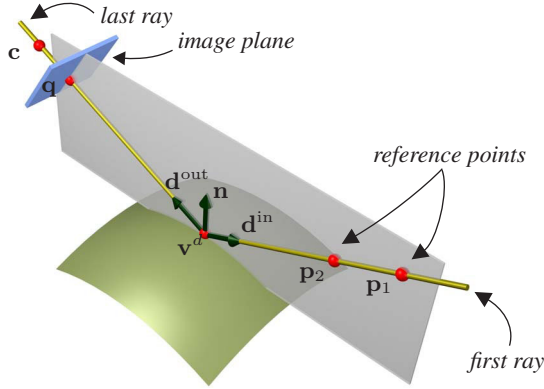


Figure 4. Visualizing the three properties of a light path. Vectors  $\mathbf{n}$ ,  $\mathbf{d}^{\text{in}}$ ,  $\mathbf{d}^{\text{out}}$  are always coplanar. In specular reflection, shown above, the angle between  $\mathbf{n}$  and  $\mathbf{d}^{\text{in}}$  is always equal to that of  $\mathbf{n}$  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 [31]. Hence, knowing one angle allows us to determine the other in both cases.

While not previously used for reconstruction, the Double-Correspondence Property has been noted in the context of environment matting [5] and camera calibration [22]. 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. This distinction is especially important in interpreting the results of our analysis.

### 3. Tractable Light-Path Triangulations

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

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

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

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

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.

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 and leads directly to a reconstruction algorithm. See [28] 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.

#### 3.1. Mirrors: One Viewpoint, Two Reference Points

**Proposition 1**  $\langle 1, 1, 2 \rangle$ -triangulation is tractable.

*Proof:* Let  $\mathbf{c}$  be a known viewpoint and let  $\mathbf{q}$  be a point on the (known) image plane whose light path has exactly one vertex  $\mathbf{v}$  (Figures 4 and 5a). Moreover, suppose that we know two distinct reference points,  $\mathbf{p}_1, \mathbf{p}_2$ , that indirectly project to  $\mathbf{q}$ . Finally, suppose that we do not know the scene’s refractive index or whether it is a mirror.

The proof follows trivially from that fact that under these conditions, both rays on the light path of  $\mathbf{q}$  are known. Specifically, the last ray is defined by the known points  $\mathbf{c}$  and  $\mathbf{q}$ . Moreover, the Double-Correspondence Property tells us that the first ray on its path passes through  $\mathbf{p}_1$  and  $\mathbf{p}_2$ . These two rays will intersect at exactly one point, which must correspond to the location of vertex  $\mathbf{v}$ . The unique depth solution is given by

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

where  $\mathbf{d}^{\text{in}}$  and  $\mathbf{d}^{\text{out}}$  are the unit vectors in the direction of the path’s two rays.<sup>4</sup> QED

While the algorithm implied by the proof of Proposition 1 is very simple, we are not aware of prior work that uses it for reconstructing specular scenes.<sup>5</sup>

#### 3.2. Glass: Three Viewpoints, Two Reference Points

**Proposition 2**  $\langle 3, 2, 2 \rangle$ -triangulation is tractable for almost all points on a generic surface with known refractive index.

<sup>4</sup>Note that if we also know that  $\mathbf{q}$ ’s light path is caused by specular reflection, the normal at  $\mathbf{q}$  is uniquely determined—it is simply the unit vector in the direction of the bisector,  $(\mathbf{d}^{\text{in}} + \mathbf{d}^{\text{out}})/2$ . 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).

<sup>5</sup>While the Double-Correspondence Property was used in [22] to recover the caustic of a mirror-based imaging system, this caustic does not coincide with the mirror’s surface and, hence, their technique is not equivalent to  $\langle 1, 1, 2 \rangle$ -triangulation.

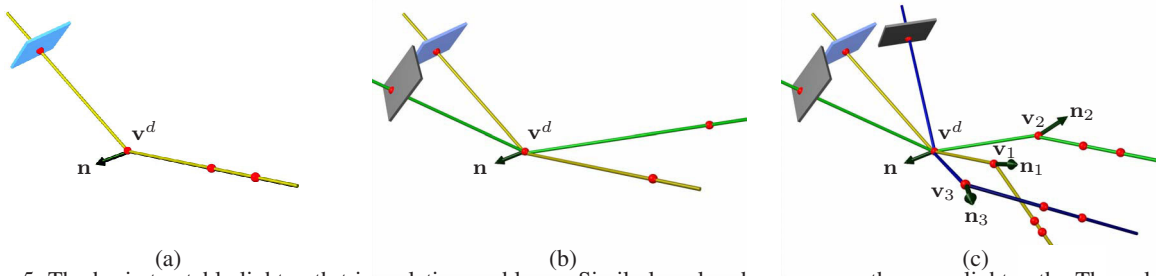


Figure 5. The basic tractable light-path triangulation problems. Similarly-colored rays are on the same light path. The unknown vertices and normals are indicated along each path. (a)  $\langle 1, 1, 2 \rangle$ -triangulation. (b)  $\langle 2, 1, 1 \rangle$ -triangulation. (c)  $\langle 3, 2, 2 \rangle$ -triangulation.

*Proof:* The proof uses two basic intuitions: (1) the set of all depth and normal assignments consistent with a single viewpoint forms a 2D “constraint surface;” 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. For concreteness, assume that the “true” light path of every image point contains two refractive vertices (Figure 5c). Paths where one or both of their vertices are reflective can be treated in an identical way.

To prove the proposition we use two facts. First, since  $M = 2$ , we know two rays on the light path of every image point (Figure 6a). Second, for scenes bounded by a generic (i.e., non-degenerate) surface [35], 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. This is because the planarity of a light path is not a stable [34] 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  $\delta$ -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. Hence, it is possible to map every pair  $(d, \delta)$  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  $\delta$  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 6b). 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  $\delta$  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  $\delta$ -values that cause rays  $l'_1$  and  $l'_2$  to intersect. To see this, note that as  $\delta$  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 [34], and since  $l'_1$  and  $l'_2$  intersect precisely at those points, it follows that for every  $d$  there is only a discrete set,  $\Delta^d$ , of  $\delta$ -values that produce a light path through  $\mathbf{q}'$ .

Finally, consider the projection,  $\mathbf{q}''$ , of  $\mathbf{v}^d$  in the third viewpoint (Figure 6c). 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  $l''_1$  and  $l''_2$  will only intersect for an isolated set of  $d$ -values. *QED*

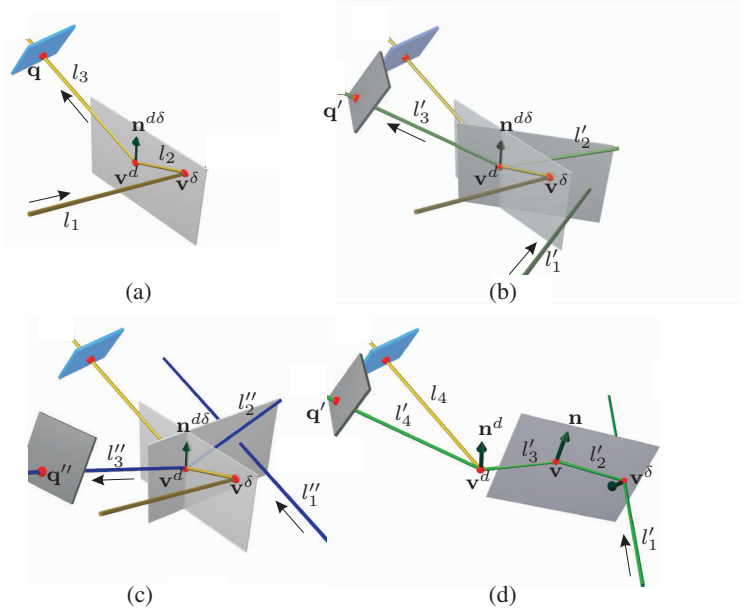


Figure 6. (a)-(c) Path geometries in proof of Proposition 2. (a) Light path of an image point  $\mathbf{q}$  in the first viewpoint. The arrow indicates the direction of incoming light. Rays  $l_1$  and  $l_3$  are known but  $l_2$  is not. The shaded plane is the plane of rays  $l_2$  and  $l_3$  and 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  $\delta$ . 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  $\delta$ , 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 3.

Beyond showing that it is possible to reconstruct general doubly-refracting and doubly-reflecting scenes, our proof

suggests a reconstruction algorithm: it tells us that we can reconstruct all four vertices and normals in the light network of a pixel  $\mathbf{q}$  by conducting a 2D search in  $(d, \delta)$ -space. The search is for a pair  $(d, \delta)$  that forces intersection both of rays  $l'_1, l'_2$  in Figure 6b and rays  $l''_1, l''_2$  in Figure 6c.

### 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 3**  $\langle N, 3, 2 \rangle$ -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 6d). 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 2. 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 6d). 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 2, the 3D position,  $\mathbf{v}^\delta$ , of this vertex can be parameterized by a single parameter  $\delta$ . Choose an arbitrary value of  $\delta$  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 4** 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.

Intuitively, Proposition 4 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 2 both as the base case and for proving the inductive step. See [36, 37] for details.

## 4. Experimental Results

We used a  $720 \times 484$ -pixel Sony DXC-9000 video camera for image acquisition and a DELL LCD display for displaying reference patterns. To calibrate the camera with respect to the plane of the LCD display, we used the Matlab Calibration Toolbox [29], 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.

**Reconstructing mirrors by  $\langle 1, 1, 2 \rangle$ -triangulation** We used the arrangement in Figures 1a and 5a. 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 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  $130 \times 230$ mm 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  $\pm 0.05$ mm, 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)). 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 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.

**Reconstructing liquids by  $\langle 2, 1, 1 \rangle$ -triangulation** See [28].

**Reconstructing glass objects by  $\langle 3, 2, 2 \rangle$ -triangulation** We used the arrangement in Figure 7 and Figure 5c. 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 display for each object rotation. The camera was tilted slightly upwards so that optical rays converging to an object point are not coplanar.

One of our test objects, a diamond-shaped glass orna-

ment, is shown in Figure 7. The object’s many planar facets, which produce complex light paths, its surface discontinuities, and its sharp, protruding tip make reconstruction especially challenging. We used five viewpoints, at  $\pm 20$ ,  $\pm 10$  and 0-degree rotations. The object extended about 4cm in depth, roughly 1m away from the camera. To reconstruct it, we used all available views and solved a  $\langle 5, 2, 2 \rangle$ -triangulation problem independently for every pixel in the 0-degree viewpoint. For each such pixel, our implementation performed a search in  $(d, \delta)$ -space for a pair of values that produce a valid light path for all viewpoints (Section 3.2 and Figures 6a-c). These values were then refined in a non-linear optimization stage. Since the light network of a pixel contains six vertices, the algorithm reconstructs six points and six normals per pixel—one on the object’s front surface and seven more on the back (Figure 3). Importantly, since we use more viewpoints than the minimum three required, the reconstruction is over-constrained and allows estimation of the object’s refractive index, which was found to be 1.53.

Figure 7 shows reconstruction results for the object’s front surface. 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. Because of this independence, normals were reconstructed accurately even near the diamond’s tip, where the surface is highly degenerate. 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 the reconstruction we measured the consistency of normals and depths within each planar facet. These quantitative measurements are shown in the table of Figure 7. They show that individually-reconstructed normals are consistent to within a few degrees, while depth measurements, which seem to produce a noisier map, show deviations on the order of 0.1 to 0.2% of the object-to-camera distance. These results, which confirm our basic theory, suggest that it is possible to recover detailed shape information for refractive objects without any knowledge of their shape, and despite the complexity of image formation.

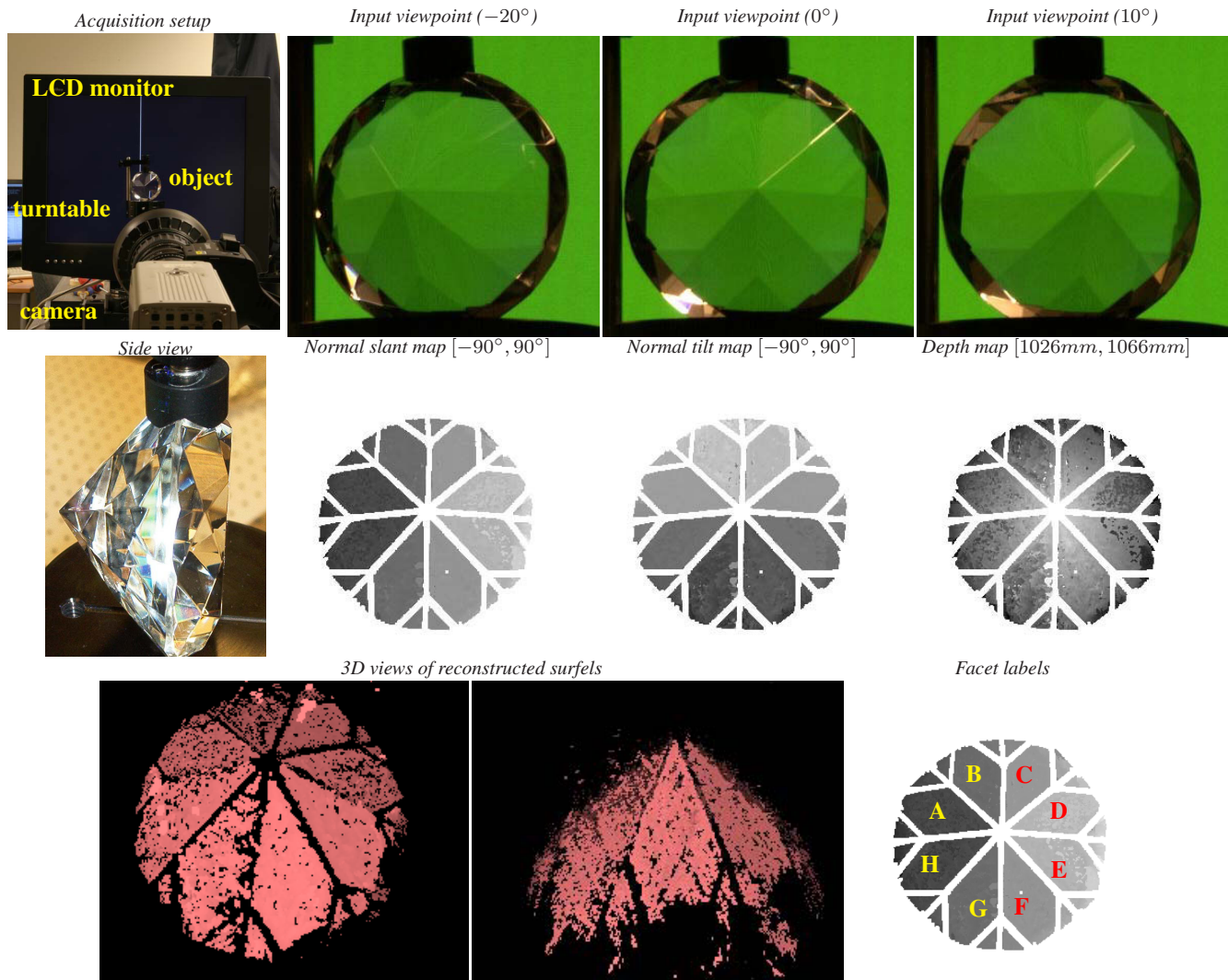
## 5. Concluding Remarks

While our 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, and (3) how to find the optimal depth at a pixel. 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.

**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 RGPIN 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,” *Proc. ECCV’98*, pp. 379–393.
- [2] A. Hertzmann and S. M. Seitz, “Shape and materials by example: A photometric stereo approach,” *Proc. CVPR’03*, pp. 533–540.
- [3] T. Zickler, P. N. Belhumeur, and D. J. Kriegman, “Helmholz stereopsis: Exploiting reciprocity for surface reconstruction,” *Proc. ECCV’02*, pp. 869–884.
- [4] W. Matusik, H. Pfister, R. Ziegler, A. Ngan, and L. McMillan, “Acquisition and rendering of transparent and refractive objects,” *Proc. 13th Eurographics Workshop on Rendering*, 2002.
- [5] D. Zongker, D. Werner, B. Curless, and D. Salesin, “Environment matting and compositing,” *Proc. SIGGRAPH’99*, pp. 205–214.
- [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. Halstead, B. Barsky, S. Klein, and R. Mandell, “Reconstructing curved surfaces from specular reflection patterns using spline surface fitting of normals,” *Proc. SIGGRAPH’96*, pp. 335–342.
- [8] J. Y. Zheng and A. Murata, “Acquiring a complete 3d model from specular motion under the illumination of circular-shaped light sources,” *IEEE T-PAMI*, v. 22, n. 8, pp. 913–920, 2000.
- [9] M. Tarini, H. P. A. Lensch, M. Goesele, and H.-P. Seidel, “3D acquisition of mirroring objects,” Tech. Rep. MPI-I-2003-4-001, Max-Planck-Institut für Informatik, 2003.
- [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*, v. 17, pp. 225–237, 1994.
- [11] S. Agarwal, S. Mallick, D. Kriegman, and S. Belongie, “On refractive optical flow,” *Proc. ECCV’04*, pp. 483–494.
- [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?,” *Proc. ICCV’03*, pp. 1025–1032.
- [15] S. Savarese and P. Perona, “Local analysis for 3d reconstruction of specular surfaces - part ii,” *Proc. ECCV’02*, pp. 759–774.
- [16] W. Keller and B. Gotwols, “Two-dimensional optical measurement of wave slope,” *Applied Optics*, v. 22, n. 22, pp. 3476–3478, 1983.
- [17] B. Jähne, J. Klinke, and S. Waas, “Imaging of short ocean wind waves: a critical theoretical review,” *JOSA-A*, v. 11, n. 8, pp. 2197–2209, 1994.
- [18] H. Murase, “Surface shape reconstruction of an undulating transparent object,” *Proc. ICCV’90*, pp. 313–317.
- [19] K. Ikeuchi, “Determining surface orientations of specular surfaces by using the photometric stereo method,” *IEEE T-PAMI*, v. 3, n. 6, pp. 661–669, 1981.
- [20] J. Wang and K. J. Dana, “A novel approach for texture shape recovery,” *Proc. ICCV’03*, pp. 1374–1380.
- [21] D. Miyazaki, M. Kagesawa, and K. Ikeuchi, “Transparent surface modeling from a pair of polarization images,” *IEEE T-PAMI*, v. 26, n. 1, pp. 73–82, 2004.
- [22] M. Grossberg and S. Nayar, “A general imaging model and a method for finding its parameters,” *Proc. ICCV’01*, v. 2, pp. 108–115.
- [23] R. Pless, “Two view discrete and differential constraints for generalized imaging systems,” *Proc. IEEE Workshop on Omnidirectional Vision*, 2002.



Facet label	A	B	C	D	E	F	G	H
Mean angle from average normal (deg)	3.48	3.35	1.68	4.85	5.62	2.78	6.33	4.44
RMS distance from best-fit plane (mm)	1.01	1.11	0.64	1.47	1.93	1.06	2.03	1.45

Figure 7. *Acquisition setup*: A linear translation stage moves the LCD in a forward/backward direction. During acquisition, the LCD displays a black background with a single horizontal or vertical stripe (see the supplementary videos in [37, 38] for examples). *Normal maps*: Gray-scale values correspond to the slant or tilt angle of each reconstructed normal. *Depth map*: Gray-scale values are mapped to the range indicated (white=near, black=far). *Surfel views*: For each pixel, we render a shiny square patch centered on the reconstructed 3D point and oriented along the reconstructed normal. *Facet measurements*: The average normal for each facet was computed by averaging the reconstructed normal for all pixels in the facet’s footprint. The best-fit plane was computed by fitting a plane to the reconstructed depths at those pixels using least squares.

- [24] A. Sanderson, L. Weiss, and S. Nayar, “Structured highlight inspection of specular surfaces,” *IEEE T-PAMI*, v. 10, n. 1, pp. 44–55, 1988.
- [25] T. Bonfort and P. Sturm, “Voxel carving for specular surfaces,” *Proc. ICCV’03*, pp. 591–596.
- [26] A. Blake, “Specular stereo,” *Proc. IJCAI’85*, pp. 973–976.
- [27] M. Oren and S. K. Nayar, “A theory of specular surface geometry,” *Proc. ICCV’95*, pp. 740–747.
- [28] N. Morris and K. N. Kutulakos, “Dynamic refraction stereo,” *Proc. ICCV’05*.
- [29] J.-Y. Bouguet, “MATLAB camera calibration toolbox.” [http://www.vision.caltech.edu/bouguetj/calib\\_doc/](http://www.vision.caltech.edu/bouguetj/calib_doc/).
- [30] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi, “Efficiently combining positions and normals for precise 3d geometry,” *Proc. SIGGRAPH’05*. To appear.
- [31] A. S. Glassner, *Principles of Digital Image Synthesis*. Morgan Kaufmann, 1995.
- [32] S. Nayar, K. Ikeuchi, T. Kanade, “Surface reflection: Physical and geometrical perspectives,” *IEEE T-PAMI*, v. 13, pp. 611–634, 1991.
- [33] J. Gluckman and S. Nayar, “Planar catadioptric stereo: geometry and calibration,” *Proc. CVPR’99*, pp. 22–28.
- [34] V. Guillemin and A. Pollack, *Differential Topology*. Prentice-Hall, 1974.
- [35] J. J. Koendering and A. J. van Doorn, “The structure of two-dimensional scalar fields with applications to vision,” *Biological Cybernetics*, v. 33, pp. 151–158, 1979.
- [36] K. N. Kutulakos and E. Steger, “A theory of specular and refractive shape by light-path triangulation,” Microsoft Research Tech. Rep. MSR-TR-2005-12.
- [37] K. N. Kutulakos, <http://www.cs.toronto.edu/~kyros/research/specular/>.
- [38] ICCV’05 Proceedings CDROM.