

# Reconstructing the Surface of Inhomogeneous Transparent Scenes by Scatter-Trace Photography

Nigel J. W. Morris

Kiriakos N. Kutulakos

University of Toronto

## Abstract

We present a new method for reconstructing the exterior surface of a complex transparent scene with inhomogeneous interior (e.g., multiple interfaces, reflective or painted interiors, etc). Our approach involves capturing images of the scene from one or more viewpoints while moving a proximal light source to a 2D or 3D set of positions. This gives a 2D (or 3D) dataset per pixel, called the *scatter trace*. The key idea of our approach is that even though light transport within a transparent scene’s interior can be exceedingly complex, the scatter trace of each pixel has a highly-constrained geometry that (1) reveals the contribution of direct surface reflection, and (2) leads to a simple “scatter-trace stereo” algorithm for computing the local geometry of the exterior surface (depth and surface normals). We present 3D reconstruction results for a variety of scenes that exhibit complex light transport phenomena.

## 1. Introduction

A major ingredient in the success of recent 3D photography algorithms is their ability to deal with *surface inhomogeneity*, i.e., to produce accurate 3D models even when a scene’s surfaces span a broad range of shape and material properties [1–4]. These algorithms apply exclusively to opaque surfaces that scatter incident light, and cannot handle scenes that contain transparent or highly-reflective media. For such scenes, the state of the art in reconstruction [5–11] is still confined to the simplest possible case—a surface bounding a single, homogeneous, transparent volume with no internal structures and no occlusion—and even this case cannot be solved without further assumptions (e.g., partially-known geometry [6, 7], a volume that causes no more than two refractions [8], or ability to immerse in a refractive-index-matched liquid [9]). Unfortunately, while objects with transparent media are very common (Figure 1), they rarely appear in isolation and rarely have a simple enough shape to fall within the realm of existing techniques. For such objects, research has concentrated on capturing their appearance rather than reconstructing them [12, 13].

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



Figure 1: Two objects used in our experiments: A hand-made solid crystal sculpture with a painted interior (about 30cm tall), and a partly-full juice bottle. The red arrow highlights the pixel used in the example of Figure 6, bottom right.

interior light transport process under these conditions has proved very difficult.

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

- **Invariance to scene interior:** To the extent possible, reconstruction performance should depend on the scene’s exterior surfaces, not the structure and complexity of media in the interior.
- **Robustness to spatially-varying reflectance and transmittance:** Reconstruction algorithms should be able to handle a wide range of surface reflectance and transmittance properties.
- **Compatibility with existing methods:** It should be possible to leverage developments in 3D photography of opaque scenes to treat issues such as noise, missing data, and occlusions.

To achieve these goals, we rely on the well-known fact that transparent scenes reflect some of the incident light, thereby behaving as partial, non-ideal specular reflectors [14, 15]. Using this as a starting point, we develop a novel technique based on *scatter-trace photography* that is specifically designed to analyze these reflections.

Scatter-trace photography involves capturing images of the scene from one or more viewpoints while moving a proximal light source to a 2D (or 3D) set of positions. This produces a 2D (or 3D) set of measurements per pixel, which we call the pixel’s *scatter trace*. Intuitively, the scatter trace of a pixel can be thought of as a “photograph” of the trajectories that light followed before interacting with the scene,

and before arriving at the given pixel (Figure 2).

The key property of the scatter trace is that direct surface reflection leaves a highly-constrained geometric “signature” in it, even when light transport within the scene’s interior is very complex. Moreover, this signature is especially prominent when the direct reflection component includes a non-negligible contribution from specular reflection. This observation leads to three main results. First, it gives rise to a geometry-based method for enhancing the contrast of the direct reflection component in each scatter trace, relative to all other modes of light transport. Second, it allows us to reduce reconstruction of inhomogeneous scenes with non-negligible specular reflectance to a generalized form of stereo matching, where we establish correspondences by comparing appropriately-processed scatter traces rather than raw pixel intensities. Third, we show that this process provides detailed information about surface orientation, at sub-pixel resolution.

Our work relies on the existence of a non-negligible specular reflection component to recover 3D shape and, as such, it is closely related to *specular stereo* methods [16–20]. These methods recover shape by analyzing the distorted appearance of patterns placed near an opaque, mirror-like scene. Of particular relevance is the work of Bonfort, Sturm and Gargallo [21] and Kutulakos and Steger [8], whose goal is to reconstruct the light path that connects each pixel in the image with two known 3D points that project to that pixel. Both approaches rely on an idealized image formation model, where light is transported along an infinitely-thin, single-bounce path corresponding to direct specular reflection off an opaque mirror. Our approach can be thought of as generalizing these methods to the case of non-ideal, inhomogeneous transparent scenes, i.e., scenes whose interior contributes significantly to appearance and whose exterior is not perfectly specular.

Although we do not explicitly decompose photos into direct, indirect, and specular components, a weaker form of separation—relative contrast enhancement—occurs implicitly as part of our generalized stereo matching procedure. In this respect, our work can also be viewed as a geometry-driven alternative to existing layer decomposition methods (direct vs. indirect [22, 23], specular vs. diffuse [25], reflected vs. transmitted [15, 26]). These methods employ a variety of tools, including active illumination [22, 23], polarization analysis [24], optical flow estimation [15], and natural image statistics [27]. Unfortunately, none of them apply to the case of general, inhomogeneous transparent scenes. For example, recent algorithms for direct/indirect separation [22, 23] break down in the presence of strong specular reflection, and polarization state stops being a robust separation cue in the presence of refraction, reflection and scattering in a scene’s unknown interior.

Our work offers four main contributions over the current state of the art. First, we derive a simple “scatter-trace stereo” algorithm for reconstructing the exteriors of scenes with transparent surfaces and inhomogeneous interiors. Second, we show that scatter-trace photography provides a natural means for revealing the scattering properties



Figure 2: *Center, back to front*: a teapot made of thin glass about 2mm-thick; a solid crystal ornament with an internal air bubble; a mug with a 2mm-thick enclosure made of clear plastic and an opaque interior cavity made of a purple specular material (see Figure 6, bottom right for another view). *Sidebars*: Scatter traces for the four pixels indicated by arrows. These pixels receive light (1) by direct reflection at an opaque point; (2) by direct reflection and by internal reflection at the back-surface of the thin teapot wall; (3) by direct reflection and by a secondary internal reflection; and (4) by direct reflection, internal reflection at the back-surface of the clear enclosure, and internal reflection off the purple interior. Note the distinct “traces” associated with each propagation mode.

of complex scenes. Third, by reducing the reconstruction problem to a simple pairwise pixel-matching criterion, our work suggests that reconstruction of inhomogeneous transparent scenes may be possible by simply replacing the “data term” in existing stereo formulations. Fourth, our results show that scatter trace analysis enables reconstruction in the presence of complex shapes and spatially-varying surface reflectance and transmittance properties. We are not aware of other image-based methods capable of reconstructing scenes of this optical complexity.

## 2. Scatter-Trace Photography

Scatter-trace photography provides a convenient way to capture the interaction of a scene with proximal point light sources and viewpoints. The most general way of “probing” this interaction is to place a point light source at some position near the scene, emit radiance only within a differential solid angle along some direction, and then measure incident radiance at some other position and direction (Figure 3a). The set of all such measurements is a ten-dimensional function that we call the *plenoptic scatter function*. This function describes how the scene scatters incident light and takes into account distance-dependent effects (e.g., that objects appear dimmer as the point light source moves farther away from them). As such, the plenoptic scatter function generalizes the familiar notions of the 4D light field [28], the 5D plenoptic function [29], and the 8D reflectance field [12, 25].

The plenoptic scatter function has extremely high dimensionality and would be very difficult to capture. Scatter-trace photography captures a specific 5D (or 4D) “slice” of

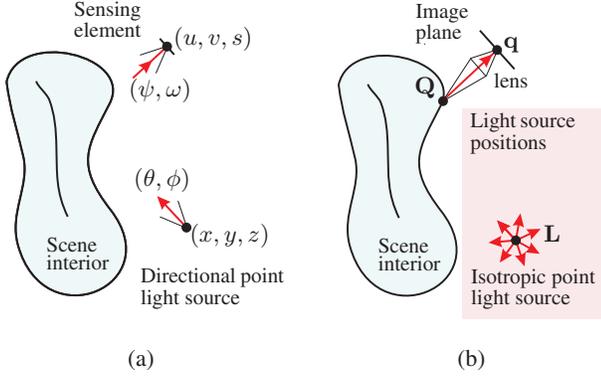


Figure 3: (a) To obtain a sample of the plenoptic scatter function we must choose a position and orientation for both the light source and the sensor (five degrees of freedom each). (b) To measure the scatter trace of pixels from a fixed viewpoint, we move the light source to every position inside a region of space (shaded).

this function that is both easy to capture and provides strong information about scene geometry. In particular, suppose that we observe the scene from a single viewpoint while illuminating it with an isotropic point light source (Figure 3b). In this case, every position  $(x, y, z)$  of the light source produces a distinct image. This image represents radiance that leaves the source equally in all directions and reaches the camera’s image plane after interacting with the scene.<sup>1</sup>

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

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

Note that if a pixel’s scatter trace is zero for some light source position, then no light passing through that position can possibly contribute radiance to that pixel. Therefore, the non-zero region of a pixel’s scatter trace is the set of all points that light can pass through to reach that pixel.

## 2.1. Direct and Indirect Scatter Traces

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

$$T_{\mathbf{q}} = T_{\mathbf{q}}^{\text{D}} + T_{\mathbf{q}}^{\text{I}}, \quad (1)$$

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

In general, the indirect component will be a significant fraction of the total scatter trace, and this fraction will vary from pixel to pixel (Figure 2). Fortunately, the direct and the

<sup>1</sup>For simplicity, assume that pixels measure radiance directly [30].

indirect components of the scatter trace have a distinct spatial structure. Here we exploit this difference to enhance the contrast of the direct component and use it for single-view reconstruction (Section 3.2) and multi-view stereo matching (Section 4).

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

**Direct reflection without scattering** Consider an infinitesimally-small pixel  $\mathbf{q}$  that is perfectly focused at a point on a planar mirror. As the lens aperture shrinks to a zero, only light sources along one incident ray  $\xi$  will contribute to the pixel’s radiance (Figure 4a). This ray is along the direction of specular reflection. We call the resulting scatter trace the *impulse scatter trace*,  $T_{\mathbf{q}}^{\xi}$ , of pixel  $\mathbf{q}$ .

For a given light source position along this ray, the radiance received at the in-focus surface point obeys a squared-distance falloff [31]. Since the radiance at  $\mathbf{q}$  cannot be larger than this radiance, the impulse scatter trace is a single “streak,” whose intensity diminishes with distance from the point of reflection.

**Generalized specular reflection** If the surface point projecting to an infinitesimally-small pixel  $\mathbf{q}$  is not a planar mirror, the pixel will receive light even from light sources that are not on a single ray. To account for this behavior, we model the micro-geometry of exterior and interior surfaces as a distribution of planar micro-facets that varies arbitrarily from point to point [32]. Furthermore, we assume that each micro-facet acts as a mirror reflector, with minimal masking or shadowing [33]. This model accurately represents the reflectance properties of smooth surfaces (e.g., glass, polished metal) and accounts for high-curvature reflectors.

Consider the surface point  $\mathbf{Q}$  that projects to pixel  $\mathbf{q}$ . We can express the point’s micro-facet distribution as a probability distribution  $D(\theta, \phi)$  over the unit sphere, with the angles  $(\theta, \phi)$  corresponding to a unique normal. In this case, the scatter trace is a weighted superposition of impulse scatter traces, one for each incoming ray (Figure 4b):

$$T_{\mathbf{q}}^{\text{D}}(\mathbf{L}) = \int_{\text{rays through } \mathbf{Q}} D(\theta^{\xi}, \phi^{\xi}) T_{\mathbf{q}}^{\xi}(\mathbf{L}) d\xi, \quad (2)$$

where the normal  $(\theta^{\xi}, \phi^{\xi})$  is the bisector of ray  $\xi$  and the visual ray through pixel  $\mathbf{q}$ .

Intuitively, as the surface at the point departs from a planar mirror, the scatter trace spreads into a “fan” of converging streaks (Figure 2, pixel 1). The point of convergence of this fan is always the surface point projecting to pixel  $\mathbf{q}$ . Moreover, the fan’s intensity decreases monotonically in a radial direction away from that point.

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

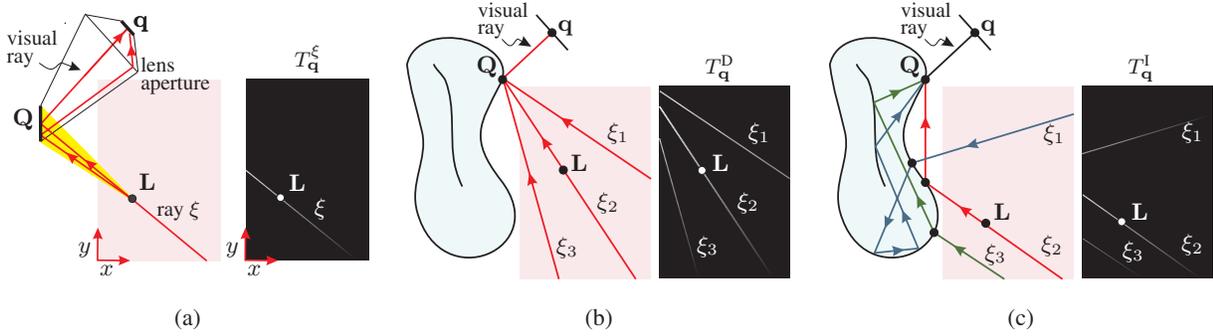


Figure 4: (a) Scatter trace due to reflection off a planar mirror. The radiance contributing to a finite-sized pixel  $q$  is transported along a bundle of rays that originate at  $q$ ; pass through the lens aperture; are specularly reflected at a small in-focus planar facet  $Q$ ; and converge at the light source position. The radiance at pixel  $q$  is therefore bounded from above by the total radiance incident at  $Q$ , i.e., the radiance transported along the blue-shaded region. In the limit, as the size of  $q$ ,  $Q$  and the aperture goes to zero, the scatter trace becomes concentrated on a single ray  $\xi$  and defines the *impulse scatter trace* along  $\xi$ . (b) Generalized specular reflection produces a “fan” of impulse scatter traces (only 3 are shown, for clarity). (c) The indirect scatter trace produced by light transport along 3 paths ending at  $q$ .

impulse scatter trace in this set represents the contribution of light that travels along an arbitrary ray  $\xi$  until it hits the object, and then follows a general piecewise-linear path to pixel  $q$  (Figure 4c):

$$T_q^I(\mathbf{L}) = \int_{\text{all rays}} \tau_q(\xi) T_q^\xi(\mathbf{L}) d\xi. \quad (3)$$

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

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

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

### 3. 3D Shape from Scatter-Trace Constraints

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

- **Viewing-ray intersection:** They must all intersect the pixel’s viewing ray.
- **Convergence:** They must converge to a single point on the viewing ray, and this point must coincide with the surface point that caused the direct reflection.
- **Monotonicity:** Their intensity must decrease monotonically in a radial direction away from the convergence point.

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

$$T_q = T_{qd}^D + T_{qd}^I. \quad (4)$$

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

#### 3.1. Estimating the Direct Scatter Trace

In the absence of additional information about the scene, the decomposition of Eq. (4) is not unique.<sup>2</sup> In light of this ambiguity, we compute the most conservative estimate of the direct component, i.e., an estimate that is guaranteed to be at least as large as that of the “true” direct scatter trace when the hypothesis corresponds to the correct depth  $d^*$ :

$$T_{qd^*}^D(\mathbf{L}) \geq T_q^D(\mathbf{L}) \quad \text{for all } \mathbf{L}. \quad (5)$$

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

In particular, each depth  $d$  defines a unique 3D convergence point,  $Q_d$ , for all “streaks” of the direct component. These streaks must therefore lie on the pencil of rays through  $Q_d$ . To enforce the convergence constraint, we rectify the pixel’s scatter trace so that all these rays become parallel to the  $x$  axis of the coordinate system. This is a linear projective warp that maps point  $Q_d$  to the point at

<sup>2</sup>For example, an estimated direct component that is zero everywhere trivially satisfies the three constraints, for any depth.

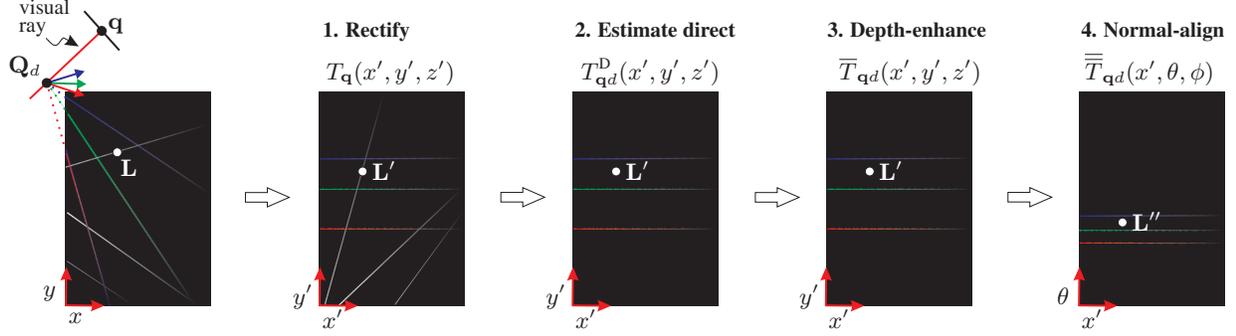


Figure 5: Basic steps in scatter-trace analysis. Each step corresponds to a transformation of the pixel’s original scatter trace, for a specific depth hypothesis. Red, green and blue colors indicate three impulse scatter traces that contribute to the pixel  $\mathbf{q}$ ’s direct scatter trace for the “true” depth  $d$ . The indirect scatter trace component is shown in white. Colored vectors in the leftmost figure are along the bisector of the viewing ray and the similarly-colored impulse scatter trace (see Section 4 for details and Figure 6, bottom right, for a real-scene example).

infinity along  $x$ . In two dimensions, it corresponds to standard epipolar image rectification [34] with the epipole at  $\mathbf{Q}_d$  (Figure 5-1).

Every line parallel to the  $x$  axis in this rectified scatter trace corresponds to a distinct ray through  $\mathbf{Q}_d$ . Geometrically, the rectification operation converts the scatter trace’s original spatial  $(x, y, z)$  coordinates into coordinates  $(x', y', z')$  that encode ray direction (coordinates  $y'$  and  $z'$ ) and position along the ray (coordinate  $x'$ ). In this coordinate system, the monotonicity constraint tells us that the intensity of the direct component must be non-increasing along the  $x'$ -axis. Since we are dealing with discrete measurements, we enforce the constraint recursively by simply computing a “running minimum” across points of the rectified scatter trace in the  $x'$ -direction (Figure 5-2):

$$T_{\mathbf{q}d}^D(\mathbf{L}') = \min(T_{\mathbf{q}}(\mathbf{L}'), T_{\mathbf{q}d}^D(\mathbf{L}' - \mathbf{X}')) , \quad (6)$$

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

Note that the computation in Eq. (6) leaves unaffected any signal that decreases monotonically along the  $x'$ -axis. As such, it will not attenuate the pixel’s “true” direct scatter trace when rectification occurs at the correct depth. Intuitively, this maximally-conservative, depth-based estimate of the direct scatter trace can be thought of as a scatter trace whose direct component is contrast-enhanced at the correct depth. This is because the “true” indirect scatter trace will typically not satisfy all three constraints (ray intersection, convergence, monotonicity) and, hence, will be attenuated for every depth hypothesis.

### 3.2. Single-View Shape from the Scatter Trace

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

In practice, we discretize the depths and evaluate the consistency between  $T_{\mathbf{q}}$  and  $T_{\mathbf{q}d}^D$  for each pixel  $\mathbf{q}$  and depth

$d$ . To do this, we first measure the point-wise consistency between these two scatter traces under the assumption of additive Gaussian noise with standard deviation  $\sigma$ ,

$$W(d, \mathbf{L}') = \exp -\frac{1}{\sigma^2} [T_{\mathbf{q}}(\mathbf{L}') - T_{\mathbf{q}d}^D(\mathbf{L}')]^2 , \quad (7)$$

and then use this consistency to enhance the direct component in the original scatter trace measurements (Figure 5-3):

$$\bar{T}_{\mathbf{q}d}(\mathbf{L}') = W(d, \mathbf{L}') T_{\mathbf{q}}(\mathbf{L}') . \quad (8)$$

In effect, this enhancement operation weighs each original scatter trace measurement by an upper-bound estimate of the likelihood that it was due to direct reflection. Our final metric aggregates the weighted measurements in Eq. (8) across all positions:

$$\mathcal{C}^1(d) = \sum_{\mathbf{L}'} \bar{T}_{\mathbf{q}d}(\mathbf{L}') . \quad (9)$$

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

## 4. Scatter-Trace Stereo

The indirect component of the scatter trace cannot be ignored when reconstructing scenes with transparent surfaces. This means that we cannot use the criterion in Eq. (9) to assess the validity of a given depth hypothesis. We therefore generalize our single-view analysis by evaluating the mutual consistency of scatter traces at corresponding pixels in two or more views.

In particular, let  $\mathbf{q}_1, \mathbf{q}_2$  be a hypothesized correspondence between two pixels in a pair of views, let  $d$  be the depth implied by this correspondence, and let  $\bar{T}_{\mathbf{q}_1d}$  and  $\bar{T}_{\mathbf{q}_2d}$  be their rectified and depth-enhanced scatter traces (Eqs. (6)-(8)). To compare these two scatter traces, we first warp them in a way that makes point-wise comparisons meaningful and then simply compute their cross-correlation:

$$\mathcal{C}^2(d) = \sum_{\mathbf{L}''} \bar{T}_{\mathbf{q}_1d}(\mathbf{L}'') \bar{T}_{\mathbf{q}_2d}(\mathbf{L}'') , \quad (10)$$

where  $\mathbf{L}''$  denotes positions in warped coordinates and  $\bar{T}_{\mathbf{q}_1 d}, \bar{T}_{\mathbf{q}_2 d}$  denote the “aligned” versions of  $\bar{T}_{\mathbf{q}_1 d}, \bar{T}_{\mathbf{q}_2 d}$ . At a superficial level, the metric in Eq. (10) can be thought of as a direct extension of traditional correlation-based stereo matching to multi-view scatter-trace photography.

The two outstanding questions are how to define the alignment warps in the two views, and under what conditions is the cross-correlation in Eq. (10) physically meaningful? We answer both questions by observing that the depth-enhanced scatter traces strongly constrain the normal(s) of the surface point  $\mathbf{Q}_d$ , projecting to the two pixels.

Every line along the  $x'$ -axis of the rectified coordinate system corresponds to a unique ray of incidence at point  $\mathbf{Q}_d$ . Given a pixel  $\mathbf{q}_i$  and such a ray, there is a unique surface normal that specularly reflects light to pixel  $\mathbf{q}_i$  from light sources on that ray (Figure 5, far left). This normal is the bisector of the ray of incidence and the visual ray through pixel  $\mathbf{q}_i$ , and can be represented by two angles,  $\theta_i(y', z')$  and  $\phi_i(y', z')$ .

A high value of the direct scatter trace for a position  $(x', y', z')$  along a specific ray of incidence can be interpreted as a “vote” for its corresponding normal. More generally, the depth-enhanced scatter trace along view  $i$  can be thought of as voting for the micro-facet distribution at point  $\mathbf{Q}_d$ . To compare two such scatter traces point by point, we align them so that points in two different scatter traces correspond to the same normal. This alignment operation transforms rectified coordinates  $(x', y', z')$  into a surface-centered,  $x'\theta\phi$ -coordinate system that is defined in terms of the surface normal at the hypothesized depth (Figure 5-4):

$$\bar{T}_{\mathbf{q}_i d}(x', \theta_i(y', z'), \phi_i(y', z')) = \bar{T}_{\mathbf{q}_i d}(x', y', z'). \quad (11)$$

In this coordinate system, the cross-correlation of Eq. (10) can be thought of as measuring the intersection of two micro-facet distributions. It will be maximized at the depth where there is maximal overlap between them, i.e., where their intersection maximally accounts for the estimated direct scatter traces in both views. The distribution itself can be computed by aggregating the votes for each normal at the depth  $d$  that maximizes Eq. (10), i.e.,

$$D(\theta, \phi) = \sum_{x'} \bar{T}_{\mathbf{q}_1 d}(x', \theta, \phi) \bar{T}_{\mathbf{q}_2 d}(x', \theta, \phi). \quad (12)$$

Therefore, maximizing Eq. (10) leads to a highly-detailed description of the local surface projecting to each pixel—both a depth  $d$  and its micro-facet distribution.

## 5. Experimental Results

We acquired 2D scatter traces using the acquisition setup in Figure 6, bottom right, with the camera about 1m away from the objects. A 2-pixel-wide vertical stripe on the LCD monitor acted as our light source, giving us 797 distinct positions that spanned a range of 41cm in the  $y$ -direction. The monitor was physically translated in the  $x$ -direction to 6 positions, spanning a range of 6cm depending on the scene.

Instead of displaying vertical stripes individually, we displayed Hadamard patterns according to Schechner *et al.*’s illumination multiplexing method [35]. Objects were placed on a turntable and rotated by  $10^\circ$  from their initial position to obtain a stereo pair. Neither the camera nor the LCD were calibrated radiometrically. For geometric calibration, we used a procedure similar to [21].

Since depth can be computed only for pixels that receive some light, pixels whose entire scatter trace was below an intensity threshold were pruned prior to reconstruction. We used a threshold equal to  $K$  times the median intensity across all scatter traces and all pixels. This “global” median can be thought of as providing an estimate of the noise level in the dataset—since each pixel receives light from only a small subset of light source positions, intensities below this median are effectively due to noise. We used  $K = 10$  for all experiments.

We assigned depth independently to each pixel in the first view by first evaluating metric  $\mathcal{C}^2(d)$  in Eq. (10) at every possible integer disparity, and then “naively” choosing the disparity that maximized this metric. Since we relied on integer disparities, depth resolution was about 2mm. To assign a surface normal, we used the peak of the pixel’s normal distribution,  $D$ , computed for that disparity. We did not apply any post-processing to the computed points and normals (e.g., smoothing, outlier rejection, etc). Our reconstruction procedure did not involve any tunable parameters.

**Fish sculpture** This scene represents an “easy” case for our method: its shape allowed a reasonably complete reconstruction because most pixels in its footprint received some illumination via direct specular reflection (Figures 1 and 6). Despite the fact that all pixels were reconstructed independently, both the depth map and the normal map are highly uniform and capture fine surface details, including high-curvature regions near the beak and the eye. To assess ground-truth reconstruction accuracy, we attached corner-like markers at nine positions spanning the object’s surface and used a hand-held 3D point probe to digitize their ground-truth 3D coordinates. RMS reconstruction error for these points was 1.4mm.

Failures of the method correspond to (1) “missing” pixels, that were not reconstructed at all because they did not receive any light from the light source, and (2) reconstructed points that deviated significantly from their “true” positions. The latter type of failure occurs at pixels near the object’s silhouette, where the surface is viewed very obliquely, and at pixels where the direct reflection component has small magnitude. These pixels are readily identifiable because the magnitude of the matching criterion at the optimal depth,  $\max_d \mathcal{C}^2(d)$ , is very low (Figure 6). Although we did not attempt to do so, it should be possible to use this magnitude as a confidence measure for outlier rejection.

**Juice bottle** The primary challenges in this scene were the rather complex geometry of the bottle’s upper section and the presence of an opaque label on the surface. Our method was able to reconstruct the label’s surface quite well, albeit with more noise compared to the regions of exposed glass. While depths and normals in those regions were re-

constructed well, the presence of self-occlusions and high curvatures meant that relatively few visible surface points reflected light toward the camera. This caused significant gaps in the reconstruction, which would require additional viewpoints to complete. In this respect, our approach is significantly less efficient than techniques for reconstructing opaque non-specular scenes, where almost all surface points visible to a pair of views can be reconstructed in one step.

**Multiple objects** This was by far the most challenging scene, with a variety of complex light transport phenomena and occlusions. Three observations can be made about our results. First, despite this complexity, we obtained detailed reconstructions for significant parts of the scene, including small high-curvature structures (e.g., a portion of the teapot lid). RMS reconstruction error in this scene was  $1.45mm$ , evaluated for five points with known ground truth on the three surfaces. Second, the specular reflectance properties of the scene’s opaque regions were not sufficient to enable their reconstruction in this case, because of the low magnitude of their scatter trace. Third, the high curvature of the interior bubble and the lack of direct reflection at pixels in its footprint led to the reconstruction of a “phantom” surface. This artifact can be mitigated by collecting additional views that cause more points on the exterior surface to reflect light directly toward the camera. More generally, however, we believe that correct treatment of interior structures should involve reasoning analogous to occlusion handling in multi-view stereo, where depth hypotheses are analyzed globally, and reconstructed front-to-back (or outside-in).

## 6. Concluding Remarks

A key contribution of our work is to show that, despite the highly-complex optical properties of inhomogeneous transparent scenes, accurate reconstruction is indeed possible with simple algorithms. Looking forward, we believe that even more general reconstruction problems are now coming within reach. We are currently investigating the reconstruction of complete surface models in the presence of occlusions, and the reconstruction of scene interiors, along with their surface.

**Acknowledgements** This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under the RGPIN program, by the U.S. National Science Foundation under Grant No. IIS-0413198, and by the Province of Ontario under the OGSST and PREA programs.

**Appendix: 2D scatter traces** Suppose we illuminate the scene with a linear light source that is oriented along the  $z$  axis and can move on the  $xy$ -plane. The radiance incident at pixel  $\mathbf{q}$  for light source position  $(x, y)$  is just the integral of the pixel’s 3D scatter traces along  $z$ :

$$T_{\mathbf{q}}(x, y) = \int T_{\mathbf{q}}(x, y, z) dz . \quad (13)$$

Now consider each of the three types of light transport discussed in Section 2.1 for the 3D case. If the point projecting

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

## References

- [1] S. J. Koppal and S. G. Narasimhan, “Clustering appearance for scene analysis,” *Proc. CVPR*, pp. 1003–1012, 2006.
- [2] J. E. Davis, R. Yang, L. Wang, “BRDF invariant stereo using light transport constancy,” *Proc. ICCV*, pp. 436–443, 2005.
- [3] T. Zickler, P. N. Bellhumeur, and D. J. Kriegman, “Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction,” *Proc. ECCV*, pp. 869–884, 2002.
- [4] T. Chen, H. Lensch, C. Fuchs, H.-P. Seidel, “Polarization and phase-shifting for 3D scanning of translucent objects,” *Proc. CVPR*, 2007.
- [5] 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, pp. 913–920, 2000.
- [6] D. Miyazaki and K. Ikeuchi, “Inverse polarization raytracing: Estimating surface shapes of transparent objects,” *Proc. CVPR*, pp. 910–917, 2005.
- [7] M. Ben-Ezra and S. Nayar, “What does motion reveal about transparency?,” *Proc. ICCV*, pp. 1025–1032, 2003.
- [8] K. N. Kutulakos and E. Steger, “A theory of refractive and specular 3D shape by light-path triangulation,” *Proc. ICCV*, pp. 1448–1455, 2005.
- [9] B. Trifonov, D. Bradley, W. Heidrich, “Tomographic reconstruction of transparent objects,” *Proc. EGSR*, pp. 51–60, 2006.
- [10] N. Morris and K. N. Kutulakos, “Dynamic refraction stereo,” *Proc. ICCV*, pp. 1573–1580, 2005.
- [11] J. Sharpe, *et al.*, “Optical projection tomography as a tool for 3d microscopy and gene expression studies,” *Science*, v. 296, pp. 541–545, 2002.
- [12] D. E. Zongker, D. M. Werner, B. Curless, and D. H. Salesin, “Environment matting and compositing,” *Proc. SIGGRAPH*, pp. 205–214, 1999.
- [13] W. Matusik, *et al.*, “Acquisition and rendering of transparent and refractive objects,” *Proc. EGSR*, pp. 267–278, 2002.
- [14] M. A. Halstead, *et al.*, “Reconstructing curved surfaces from specular reflection patterns using spline surface fitting of normals,” *Proc. SIGGRAPH*, pp. 335–342, 1996.
- [15] R. Szeliski, S. Avidan, and P. Anandan, “Layer extraction from multiple images containing reflections and transparency,” *Proc. CVPR*, pp. 246–253, 2000.
- [16] Marco Tarini, Hendrik P. A. Lensch, Michael Goesele, and Hans-Peter Seidel, “3D acquisition of mirroring objects using striped patterns,” *Graphical Models*, v. 67, n. 4, 2005.
- [17] S. Savarese, M. Chen, and P. Perona, “Local shape from mirror reflections,” *IJCV*, v. 64, n. 1, pp. 31–67, 2005.
- [18] A. C. Sanderson, L. E. Weiss, and S. K. Nayar, “Structured high-light inspection of specular surfaces,” *IEEE T-PAMI*, v. 10, pp. 44–55, 1988.
- [19] M. Oren and S. Nayar, “A theory of specular surface geometry,” *IJCV*, v. 24, n. 2, pp. 105–124, 1996.
- [20] T. Bonfort and P. Sturm, “Voxel carving for specular surfaces,” *Proc. ICCV*, pp. 591–596, 2003.
- [21] T. Bonfort, P. Sturm, and P. Gargallo, “General specular surface triangulation,” *Proc. ACCV*, pp. 872–881, 2006.
- [22] S. K. Nayar, G. Krishnan, M. D. Grossberg, and R. Raskar, “Fast separation of direct and global components of a scene using high frequency illumination,” *Proc. SIGGRAPH*, pp. 935–944, 2006.
- [23] S. M. Seitz, Y. Matsushita, and K. N. Kutulakos, “A theory of inverse light transport,” *Proc. ICCV*, pp. 1440–1447, 2005.
- [24] K. Koshikawa, “A polarimetric approach to shape understanding of glossy objects,” in *Physics-Based Vision: Principles and Practice*, v. 1, AK Peters, 1992.

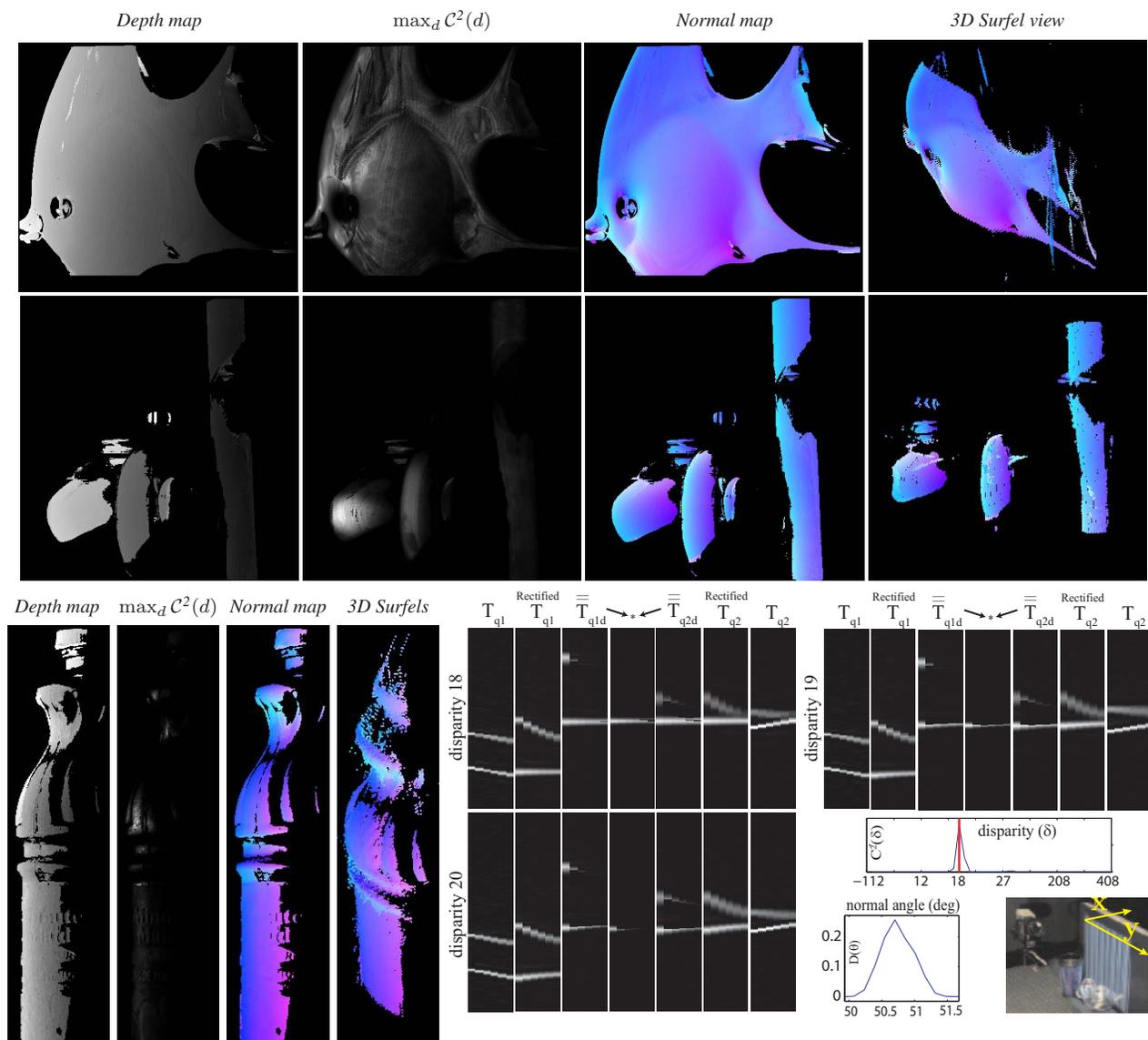


Figure 6: *Top rows and bottom-left*: Reconstruction results for the scenes in Figures 1 and 2. *Bottom right*: Scatter traces and intermediate results of applying the scatter trace stereo algorithm to the highlighted pixel in Figure 1(left). Scatter trace resolution was  $6 \times 797$  and coordinate axes are as in Figure 5. Note the double streaks, corresponding to a direct reflection component and a secondary component due to indirect reflection (this occurs for almost all pixels in this scene). Also shown are the values of the matching criterion across disparities and the recovered normal distribution (only one angle is shown). Note the unambiguous peak at the true disparity,  $\delta = 18$ , despite the presence of strong secondary illumination.

[25] P. Debevec, *et al.*, “Acquiring the reflectance field of the human face,” in *Proc. SIGGRAPH*, pp. 145–156, 2000.

[26] T. Oo, H. Kawasaki, Y. Ohsawa, and K. Ikeuchi, “Separation of reflection and transparency using epipolar plane image analysis,” *Proc. ACCV*, pp. 908–917, 2006.

[27] A. Levin, A. Zomet, and Y. Weiss, “Separating reflections from a single image using local features,” *Proc. CVPR*, pp. 306–313, 2004.

[28] M. Levoy and P. Hanrahan, “Light field rendering,” *Proc. SIGGRAPH*, pp. 31–42, 1996.

[29] E. H. Adelson and J. R. Bergen, “The plenoptic function and the elements of early vision,” in *Computation Models of Visual Processing*, MIT Press, 1991.

[30] P. E. Debevec and J. Malik, “Recovering high dynamic range radiance maps from photographs,” *Proc. SIGGRAPH*, pp. 369–378, 1997.

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

[32] D. Edwards, *et al.*, “The halway vector disk for brdf modeling,” *ACM TOG* v. 25, n. 1, pp. 1–18, 2006.

[33] S. K. Nayar, K. Ikeuchi, and T. Kanade, “Surface reflection: Physical and geometrical perspectives,” *IEEE T-PAMI* v. 13, pp. 611–634, 1991.

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

[35] Y. Y. Schechner, S. K. Nayar, and P. N. Belhumeur, “A theory of multiplexed illumination,” *Proc. ICCV*, pp. 808–815, 2003.