

3D Reconstruction of Transparent Objects with Position-Normal Consistency

Yiming Qian
 University of Alberta
 yqian3@ualberta.ca

Minglun Gong
 Memorial University of Newfoundland
 gong@cs.mun.ca

Yee-Hong Yang
 University of Alberta
 yang@cs.ualberta.ca

Abstract

Estimating the shape of transparent and refractive objects is one of the few open problems in 3D reconstruction. Under the assumption that the rays refract only twice when traveling through the object, we present the first approach to simultaneously reconstructing the 3D positions and normals of the object’s surface at both refraction locations. Our acquisition setup requires only two cameras and one monitor, which serves as the light source. After acquiring the ray-ray correspondences between each camera and the monitor, we solve an optimization function which enforces a new position-normal consistency constraint. That is, the 3D positions of surface points shall agree with the normals required to refract the rays under Snell’s law. Experimental results using both synthetic and real data demonstrate the robustness and accuracy of the proposed approach.

1. Introduction

3D reconstruction of real objects is an important topic in both computer vision and graphics. Many techniques have been proposed for capturing the shapes of opaque objects using either active [36] or passive [5] manners. However, accurately reconstructing transparent objects, made of glass and crystal, remains an open and challenging problem. The difficulties are caused by several factors. Firstly, these objects do not have their own colors but acquire their appearances from surrounding diffuse objects. Hence, conventional color/textured matching based approaches cannot be applied. Secondly, transparent objects interact with light in complex manners including reflection, refraction, scattering and absorption. Tracing the poly-linear light paths is very difficult, if not impossible. Thirdly, the behavior of refraction depends on the object’s refractive index, which is usually unknown.

Previous approaches of 3D transparent object reconstruction can be roughly classified into three groups [17, 18]: reflection-based, refraction-based, and intrusive methods. The first group attempts to reconstruct the objects by utilizing the specular highlights on the object surface

[21, 25]. By analyzing only the surface reflection properties, such approaches can reconstruct transparent objects with complex and inhomogeneous interior. However, unlike opaque objects, only a small amount of light is reflected from the transparent object’s surface. To measure the weak reflection, precise controlling and adjusting the light positions are usually required. The second group exploits the refraction characteristics of transparent objects. Many methods simplify the problem by considering only one-refraction events, either assuming that the surface facing away from the camera is planar [29] or that the object is thin [34]. As well, the refractive index is required to be known for surface normal estimation. Although the problem of two-refraction events has been investigated theoretically [20], the setup requires high precision movements of both the object and the light source, making the approach hard to use and the results difficult, if not impossible, to reproduce. Finally, intrusive methods either rely on special devices (*e.g.* light field probes [34], diffuse coating [12]) or by immersing the object in special liquids [14, 15, 31], which are often impractical and may even damage the objects. Hence, there is a need for a practical approach that can accurately reconstruct transparent objects using a portable setup.

This paper presents a new refraction-based approach for reconstructing homogeneous transparent objects, through which light is redirected twice. As is commonly done, inter-reflections within the object are assumed to be negligible. By introducing a novel position-normal consistency constraint, an optimization procedure is designed, which jointly reconstructs the 3D positions and normals at two refraction locations. The refractive index of the object can also be reliably estimated by minimizing a new reconstruction error metric. Further, our acquisition setup is simple and inexpensive, which consists of two cameras and one monitor, all of which do not require precise positioning.

2. Related Work

Although the problems of reconstructing dynamic wavefronts [9, 24] and gas flows [4, 19, 22] are also related to our work, here we focus our review on recent works for

static reflective and refractive surfaces, for which our approach is designed. In addition, since our approach is non-intrusive, the intrusive ones [10, 16, 31] are not discussed here. Readers are referred to the comprehensive surveys of the field [17, 18].

Shape from reflection based methods utilize the specular property of the surface. Such methods bypass the complex interactions of light with the object as it travels through the object by acquiring the linear reflectance field. Hence, inhomogeneous transparent objects can be reconstructed. Tarini *et al.* [30] acquire light reflections of mirror object against a number of known background patterns and then alternately optimize the depths and normals from reflective distortions. Morris and Kutulakos [25] reconstruct complex inhomogeneous objects by capturing exterior specular highlights on the object surface. Their approach requires delicate movements of light sources and imprecise movements often introduce errors to the results. Yeung *et al.* [38] introduce a low-cost solution by analyzing specular highlights, which can only obtain the normal map of an object. Recently, Liu *et al.* [21] apply a frequency-based approach to establish accurate reflective correspondences, but only sparse 3D points are obtained. A common issue of the above reflection-based approaches is that the reflectance field is often corrupted by the indirect light transport within the object and various constraints are proposed to tackle it.

Refraction-based reconstruction methods rely on the refracted light, which is stronger than the reflected light for truly transparent objects and conveys unique characteristics of the objects. Many works [2, 9, 24, 26] mainly focus on reconstructing water surfaces. Typically, a known pattern is placed beneath a water tank and an optical flow algorithm is applied to obtain the correspondences between the camera pixels and the contributing points from the pattern. Pixel-point correspondence based approaches are known to have ambiguities: the 3D object point can be located at any position along the ray that corresponds to an object pixel. Hence, 3D reconstruction is usually performed with an optimization procedure and additional constraints are often required to resolve ambiguities [24, 26]. However, the accuracy of optical flow tracking often affects the reconstruction results.

Ben-Ezra and Nayar [6] develop a structure-from-motion based method for reconstructing the full 3D model of a transparent object, where the object is assumed to have a known parametric form. Wetzstein *et al.* [34] acquire the correspondences between the incident and exit rays, *i.e.* ray-ray correspondences, from a single image using light field probes and compute refraction positions through triangulation. However, their method assumes that the incident light is redirected once and thus, can only work for thin objects. Similarly, several methods [23, 29, 32] simplify the problem by focusing on only one-refraction events. In particu-

lar, they either assume that part of the surface information (*e.g.* normal) is known or that one of the refraction surfaces is known.

Kutulakos and Steger [20] categorize the refraction-based approaches based on the number of reflections or refractions involved, and discuss the feasibility of reconstruction of different cases. They show that at least three views are required to reconstruct a solid object where light is redirected twice. Our approach also handles two-refraction cases, but differs from theirs in the following aspects: i) Their approach triangulates individual light paths separately to reconstruct the corresponding surface points, whereas we use an optimization procedure to solve all points in conjunction; ii) our approach reconstructs both refraction surfaces, whereas theirs only deals with a single surface; iii) we simultaneously recover both the 3D positions and normals of the refraction surface, whereas their approach computes the surface normals based on Snell’s law in post-processing, which may not be consistent with the local shape; iv) we use only two cameras for data acquisition, while they use five, leading to more data to be captured; and v) precise object rotation and monitor translation are required in their setup and hence, applying their technique can be difficult. In contrast, our approach does not require precise positioning of the monitor, while the object is fixed during acquisition.

In addition, there are a number of other works on imaging under refraction that are worthy to mention but are outside the scope of this paper. For example, environment matting [8, 35, 41] aims to composite transparent objects into new background. Underwater imaging [3] studies recovering undistorted scenes from a sequence of distorted images. The aperture problem [37] reveals the refractive motion of transparent elements.

3. Proposed Approach

3.1. Acquisition Setup and Procedure

Our approach requires the acquisition of ray-ray correspondences before and after refraction. That is, for each observed ray refracted by the transparent object, we like to know the corresponding incident ray. As shown in Fig. 1, we use an LED monitor as the light source. Through displaying predesigned patterns on the monitor, the location of the emitting source for each captured ray can be found at pixel level accuracy. Adjusting the monitor location and repeating the process gives us two positions of the incident ray and hence, the ray direction can be determined. The same procedure is performed for the second camera, which observes the object in the opposite side of the first one.

Fig. 2 further illustrates the acquisition process in 2D. Two cameras are placed on the opposite sides of the object with their positions fixed during acquisition. For simplicity, we here refer the object surfaces on these two sides as the

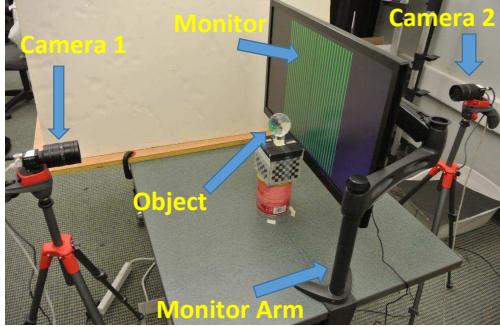


Figure 1. Our data acquisition setup, where two cameras roughly face each other. Camera 1 is capturing data in this photo. Once Camera 1 is done, the monitor is moved to the other side of the object to serve as the light source for Camera 2.

front and back surfaces, respectively. We first use Camera 1 to capture the front surface with the monitor positioned at plane m_1 . The environment matting (EM) algorithm [27] is applied to locate the contributing sources p_i on the monitor at pixel accuracy, which is achieved by projecting a set of frequency-based patterns. The monitor is then moved to plane m'_1 and the EM method is repeated. Connecting point p_i and p'_i gives us the incident ray direction \vec{d}_i^{in} for the light source. The corresponding exit ray direction \vec{d}_i^{out} is obtained from the intrinsic camera matrix, which is calibrated beforehand. We then capture the back surface using Camera 2 in a similar fashion with the monitor positioned at plane m_2 and m'_2 .

Please note that precise monitor movement is not required in our setup. The monitor can be moved by any distance and its position can be easily calibrated by displaying a checkerboard pattern [39]. It is noteworthy that instead of determining the incident ray using two monitor locations, light field probes [33] can also be used. However, we choose the monitor approach for two reasons: i) The light source locations can be determined at pixel-level accuracy and, ii) by displaying patterns with a primary color, our approach is robust to dispersion effects, whereas approaches relying on color-calibration are not.

So far, we have obtained the ray-ray correspondences w.r.t. the front and back surfaces using two cameras. In the subsections below, we present a novel reconstruction scheme that solves the following problem: *Given the dense ray-ray correspondences $(p, \vec{d}^{in}) \Leftrightarrow (c, \vec{d}^{out})$ of two cameras, how to compute the 3D positions and normals of the front and back surface points?*

3.2. Position-Normal Consistency

The seminal work [20] has shown that three or more views are required to reconstruct a single surface where the light path is redirected twice. Here we show that, by assuming the object surface is piecewise smooth, we can solve

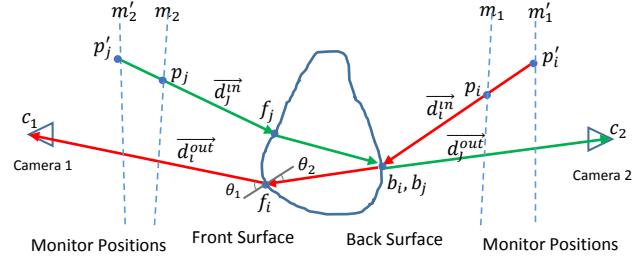


Figure 2. Our acquisition setup using a pair of cameras and one monitor as light source. Note that the monitor is moved to different positions during acquisition.

both the front and back surfaces using data captured from only two cameras. The key idea is that, for each reconstructed 3D surface point, its normal estimated based on its neighboring points should agree with the normal required for generating the observed light refraction effect.

We first explain how to measure position-normal consistency error for a given shape hypothesis. Here the object shape is represented using depth maps of the front and back surfaces, where the depth of a surface point is measured as its distance to the camera center along the camera's optical axis. Taking Camera 1 for example, as shown in Fig. 2, given a ray-ray correspondence $(p_i, \vec{d}_i^{in}) \Leftrightarrow (c_1, \vec{d}_i^{out})$ from Camera 1, the locations at which the incident ray (p_i, \vec{d}_i^{in}) and the exit ray (c_1, \vec{d}_i^{out}) meet the object are denoted as b_i and f_i , respectively. f_i can be computed from the corresponding depth map and how to compute b_i is discussed in Sec. 3.2.2. Connecting b_i and f_i gives us a hypothesis path that the ray travels through while inside the object. Hence, based on Snell's law, we can compute the normal at f_i , which is referred as the Snell normal. Furthermore, using the 3D locations of nearby points of f_i , we can also estimate the normal of f_i using Principal Component Analysis (PCA) [28], which is referred as the PCA normal. Ideally, the PCA normal and the Snell normal for the same point are the same. Hence, for the i th ray-ray correspondence, its position-normal consistency error is measured as:

$$E_{pnc}(i) = 1 - |P(i) \cdot S(i)|, \quad (1)$$

where $P(i)$ (or $S(i)$) computes the PCA (or Snell) normal at the 3D location where the exit ray (c, \vec{d}_i^{out}) leaves the object. Note that the same definition also applied to ray-ray correspondences found by Camera 2.

In addition, based on the assumption that the object surface is piecewise smooth, we also want to minimize the depth variation in each depth map D . Hence, the second error term is defined as:

$$E_{so}(D) = \sum_{s \in D} \sum_{t \in N(s)} (D(s) - D(t))^2, \quad (2)$$

where $\mathcal{N}(s)$ denotes the local neighborhood of pixel s in a given depth map D .

Combining both terms and summing over both front and back surfaces gives us the objective function:

$$\min_{D_f, D_b} \left(\sum_{i \in \Omega} E_{pnc}(i) + \lambda(E_{so}(D_f) + E_{so}(D_b)) \right), \quad (3)$$

where D_f and D_b are the depth maps for the front and the back surfaces, respectively, and Ω is the set containing all the ray-ray correspondences found by both cameras. Hence, Eq.(3) optimizes all the points in both depth maps at the same time using all available correspondence information. λ is a parameter balancing E_{pnc} and E_{so} .

In the following subsections, we present how to compute the PCA and the Snell normals under the depth map hypotheses D_f and D_b . We use the front surface observed by Camera 1 to illustrate our approach, and the back surface is processed in the same fashion.

3.2.1 Normals from Positions by PCA

Given the positions of 3D points, previous work [28] has shown that the normal of each point can be estimated by performing a PCA operation, *i.e.*, analyzing the eigenvalues and eigenvectors of a covariance matrix assembled from neighboring points of the query point. Specifically, the covariance matrix \mathcal{C} is constructed as follows:

$$\mathcal{C} = \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} (f_j - f_i)(f_j - f_i)^T. \quad (4)$$

In our implementation, we use a 5×5 neighborhood of each pixel. The PCA normal is therefore the eigenvector of \mathcal{C} with the minimal eigenvalue.

3.2.2 Normals by Snell's Law

As shown in Fig. 2, to obtain the Snell normal of f_i , the refractive index of the object, the interior ray path $\overrightarrow{b_i f_i}$ and the exit ray $\overrightarrow{f_i c_1} = \overrightarrow{d_i^{out}}$ are required. Here the refractive index is assumed to be known (how to handle objects with unknown refractive index will be discussed in Sec. 3.3). Since the front point position f_i is given, the interior ray direction $\overrightarrow{b_i f_i}$ can be obtained by locating the corresponding back point b_i . Note that b_i is observed by Camera 2 and on the line $(p_i, \overrightarrow{d_i^{in}})$. Hence, the problem of estimating the Snell normal of f_i is reduced to the problem of locating the first-order intersection between the back surface and the line $(p_i, \overrightarrow{d_i^{in}})$. A similar problem has been studied in image-based rendering, where the closest intersection between a ray and a disparity map needs to be computed. Here we apply the solution proposed in [13], which converts the 3D intersection calculation problem into a problem of finding the zero crossing of a distance function.

After getting b_i , Snell's law is applied to estimate the Snell normal $S_1(i)$ for f_i . Denote η_1 and η_2 as the refractive index of air and the object, respectively, we have $\eta_1 \sin \theta_1 = \eta_2 \sin \theta_2$, where θ_1 and θ_2 are the angles between the normal and each of the light paths, as shown in Fig. 2. Let $\Delta\theta = \theta_1 - \theta_2 = \cos^{-1}(\overrightarrow{f_i c_1} \cdot \overrightarrow{b_i f_i})$, we have:

$$\theta_1 = \tan^{-1} \left(\frac{\eta \sin \Delta\theta}{\eta \cos \Delta\theta - 1} \right), \quad (5)$$

where $\eta = \eta_2/\eta_1$ is the relative index of refraction. The Snell normal is obtained by rotating $\overrightarrow{f_i c_1}$ by angle θ_1 on the plane spanned by $\overrightarrow{f_i c_1}$ and $\overrightarrow{b_i f_i}$, that is:

$$S_1(i) = \mathcal{R}(\theta_1, \overrightarrow{f_i c_1} \times \overrightarrow{b_i f_i}) \overrightarrow{f_i c_1}, \quad (6)$$

where $\mathcal{R}(\theta, \vec{v})$ is the Rodrigues rotation matrix defined by θ and the rotation axis \vec{v} .

3.3. Optimize Depth Maps and Refractive Index

The aforementioned procedure returns the position-normal consistent model under a given refractive index hypothesis. For objects with unknown refractive indices, additional work needs to be done to estimate the proper index values. Similar to previous approaches [24, 29], our strategy here is to enumerate different refractive index values, evaluate the resulting models, and pick the best solution. However, unlike [24, 29], where the objective function to be optimized is directly used for evaluating the model quality, here a different reconstruction error metric is used.

As shown in Fig. 2, a given point b_i on the object surface may be involved in the ray-ray correspondence $(p_i, \overrightarrow{d_i^{in}}) \Leftrightarrow (c_1, \overrightarrow{d_i^{out}})$ from Camera 1 and the correspondence $(p_j, \overrightarrow{d_j^{in}}) \Leftrightarrow (c_2, \overrightarrow{d_j^{out}})$ from Camera 2. In the first ray path, b_i is the location where the incident ray enters the object, whereas in the second ray path, b_i is the location where the exit ray leaves. Two Snell normals can be computed from the two ray paths. Since the two Snell normals should be the same under the ground-truth model and the true refractive index value, their difference is a good measure of the reconstruction result. Hence, we define the reconstruction error for model D as:

$$RE(D) = \sum_{s \in \Psi} (1 - |S_b(s) \cdot S_f(s)|), \quad (7)$$

where $S_f(s)$ and $S_b(s)$ refer to the Snell normals computed using rays entering and exiting location s , respectively. Ψ is a set containing all locations on object surface that are involved in two ray-ray correspondences. It is worth noting that Eq.(3) only uses Snell normals computed using exit rays. Hence, Eq.(7) evaluate different errors as the objective function does.

Following [29], a coarse-to-fine optimization scheme is used for searching both the optimal refractive index and the

optimal depth maps. In our implementation, we first downsample the obtained correspondences to 1/4 of the original resolution, enumerate the refractive index in the range of [1.2, 2.0] with increments of 0.05, and compute the optimal shape under each index value by minimizing Eq.(3). The relative index with the minimal reconstruction error as defined in Eq.(7) is then selected to compute the final model using the full resolution ray-ray correspondences.

Optimizing Eq.(3) is difficult because of the complex operations involved in the PCA and Snell normal computations. To avoid trivial local minima, we place a checkerboard in front of the front surface and the back surface, respectively. By calibrating the checkerboards, the depth searching ranges for D_f and D_b are obtained. Now Eq.(3) becomes a bounded constrained problem. We use the L-BFGS-B method [40] to solve Eq.(3) with numerical differentiation applied.

4. Experiments

The presented algorithm is tested on both synthetic and real data. The factor λ is fixed at 50 units in the synthetic data and 0.005 mm in the real experiments. Since the PCA and Snell normal calculations for different pixels can be independently performed, they are computed in parallel. We implemented our parallel algorithm in MATLAB R2014b. Running on an 8-core PC with 3.4GHz Intel Core i7 CPU and 24GB RAM, the processing time needed for the models shown below varies between 1-2 hours.

4.1. Synthetic Object

We start with validating our approach on a synthetic sphere, where the ray-ray correspondences are generated by a ray-tracer. Specifically, the sphere is centered at $(0, 0, 2)$ with radius = 0.2. Two cameras are placed at $(0, 0, 0)$ and $(0, 0, 4)$. One observes the front surface and the other the back surface. By tracing along the poly-linear light paths, we can mathematically compute both the ground-truth positions and normals of the front and back surface points.

To evaluate the accuracy and robustness of our approach under different levels of data acquisition noise, we add zero-mean Gaussian noise to the obtained ray-ray correspondences. That is, for a given observed ray $(c, \overrightarrow{d^{out}})$, the corresponding light source locations under two monitor settings, p and p' , are both corrupted with noise of standard deviation σ ($\sigma \leq 10$ pixels). The cameras are assumed to be calibrated, *i.e.*, their locations, orientations and internal parameters are not corrupted. We evaluate the reconstruction accuracy using three measures: the root mean square error (RMSE) between the ground-truth depths and the estimated ones, the average angular difference (AAD) between the true normals and the reconstructed PCA normals, and the AAD between the true and the computed Snell normals. As shown in Fig. 3, our approach achieves high accuracy on

both position and normal estimation and is robust to varying noise level. Fig. 4 visually compares the ground truth and the reconstructed results.

In addition to simultaneously reconstructing the 3D positions and normals, our approach can estimate the refractive index of the object. Here we evaluate the stability of refractive index estimation. By assigning different relative indices η , we capture the ray-ray correspondences using our ray-tracer. Then Gaussian noise with $\sigma = 5$ is added. Fig. 5 shows the variation of reconstruction error Eq.(7) with hypothesized refractive index. It shows that the index that corresponds to the minimum of the reconstruction error is close to the true index. This means that the proposed error term Eq.(7) can effectively estimate the refractive index. In comparison, directly using the objective function Eq.(3) cannot estimate the refractive index well.

4.2. Real Refractive Objects

Three transparent objects, a Swarovski ornament, a glass ball and a green bird, are used for evaluating the proposed approach on real captured data. The “ornament” and “ball” objects have apparent dispersion effects. To properly handle that, we use two Point Grey Blackfly monochromatic cameras so that artifacts of the Bayer mosaic can be avoided. An LG IPS monitor is used to display frequency-based patterns using a single color channel [27] with a resolution of 1024×1024 . Calibration between the two cameras is challenging because they are facing each other (see Fig. 1). To address this difficulty, we place an additional camera between them and conduct pairwise camera calibrations. After calibration, the third camera is removed.

As shown in Table 1, the ornament and the glass ball each have many planar facets on its surfaces, resulting in complex light-object interaction and normal discontinuities. Fig. 6 shows our reconstruction results including the point clouds and depth maps, which are visually encouraging for both objects. More importantly, since our approach jointly optimizes the 3D positions and normals, the reconstructed normals are reconciled with the estimated point clouds.

Following [14, 20], to quantitatively assess the reconstruction accuracy, we manually label several facets shown in Table 1. For each facet, we fit a plane using the RANSAC algorithm [11]. Two measures are used to evaluate each facet: the AAD between the reconstructed normals and the fitted plane normal, as well as the mean distance error from the estimated 3D points to the plane. The quantitative measurements in Table 1 imply that the reconstructed normals and positions within each planar facet are consistent. This suggests that our approach can accurately reconstruct the piecewise planar structure without any prior knowledge of the shapes or parametric form assumptions.

Fig. 7 shows the reconstruction results of the bird. The model contains three largely separated parts. To avoid the

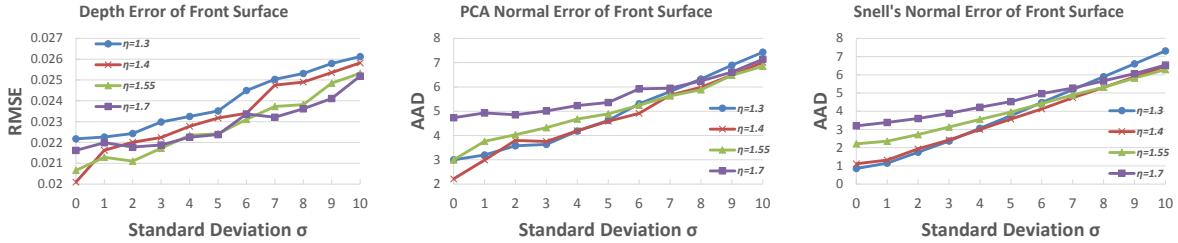


Figure 3. Reconstruction accuracy for the front surface as a function of Gaussian noise level on the synthetic sphere under different refractive indices. The plots for the back surface are similar due to the symmetric setup. 50 trials are performed under each setting.

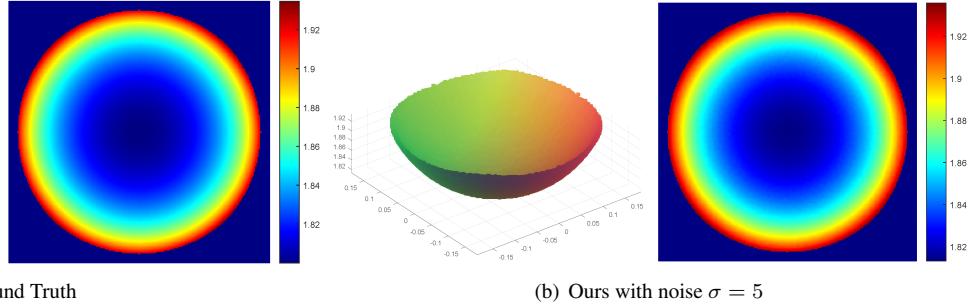


Figure 4. Visual comparisons between the ground truth and our result for the synthetic sphere. In each case, we show the point cloud for the front surface colored with the corresponding PCA normal map, as well as the depth map. The Snell normals are not shown here since they are similar to the PCA normals. Please see the supplemental materials [1] for the results of back surfaces and under other noise levels.

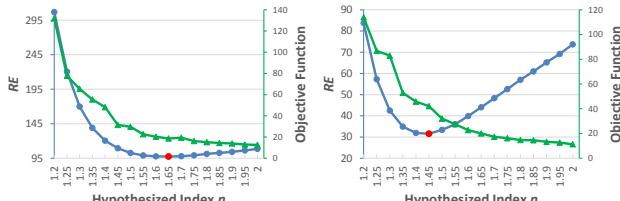


Figure 8. Refractive index estimation for the ‘‘ornament’’ (left) and ‘‘ball’’ (right) object. Blue curves plot the proposed reconstruction error term as a function of hypothesized refractive index. Green curves plot the corresponding objective function. The minimum of the blue curve is marked in red. Please see the supplemental materials [1] for the curves of the ‘‘bird’’ object.

inter-reflections between the three parts, we use tapes to block lights from the two smaller birds on the side when capturing the data. Our results successfully captures the front and back shape of the bird in the center. Note that since the bird only transmit green light, approaches relying on light field probes won’t work.

Fig. 8 shows the reconstruction error Eq.(7) under different hypothesized refractive indices. The estimated refractive index for ‘‘ornament’’ is 1.65, which agrees with the available report¹ stating that the refractive index of Swarovski crystal is between 1.5 and 1.7.

¹<http://www.crystalandglassbeads.com/blog/2012/diamonds-cubic-zirconia-swarovski-whats-the-difference.html>

5. Conclusions and Limitations

We have presented a refraction-based approach for reconstructing transparent objects. We first develop a simple acquisition setup which uses a pair of cameras and one monitor. Compared to existing methods, our system is non-intrusive, and requires no special devices or precise light source movement. By introducing a novel position-normal consistency constraint, we propose an optimization framework which can simultaneously reconstruct the 3D positions and normals of both the front and back surfaces. Note that many existing methods can only reconstruct either the depth or the normal of a single surface. In addition, we show that it is possible to estimate the refractive index of transparent objects using only two views.

Our approach works under the following assumptions: i) the object is solid and homogeneous, ii) the light path between the source and the camera goes through two refractions, and iii) the object surface is smooth enough so that surface normals can be reliably estimated using available sample points. Note that these assumptions are commonly used by refraction-based methods [20, 32]. Moreover, our acquisition is simple and inexpensive, but at the cost of capturing thousands of images since the ray-light source correspondences are required at each of the four monitor positions. As discussed above, replacing the monitor with light field probes [33] helps to reduce the number of images needed, but at the expense of loosing sampling density

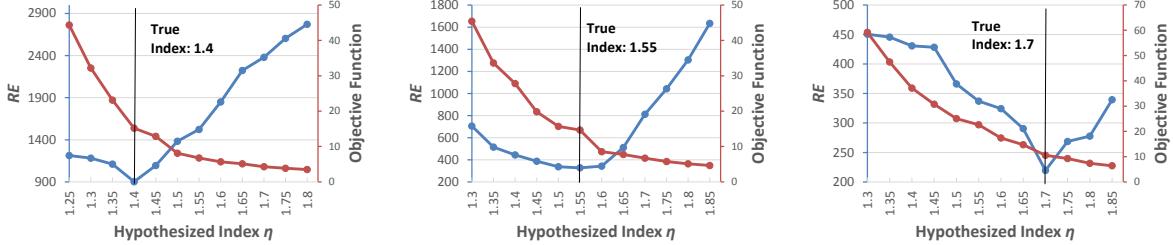


Figure 5. Refractive index estimation for the synthetic sphere. Blue curves plot the reconstruction error Eq.(7) as a function of hypothesized refractive index. Red curves plot the corresponding objective function Eq.(3). Ground-truth indices are shown with vertical lines.

		Facet	1	2	3	4	5	6
(a)		Mean positional error (mm)	0.13	0.15	0.14	0.13	0.21	0.18
		AAD of PCA normals (degree)	3.28	5.10	4.77	3.32	6.50	5.10
		AAD of Snell normals (degree)	4.27	6.07	5.63	3.93	7.12	6.02
		RANSAC position inliers (%)	50.94	40.16	41.18	49.32	39.21	44.67
(b)		Mean positional error (mm)	0.13	0.11	0.13	0.12	0.15	0.15
		AAD of PCA normals (degree)	3.83	3.10	3.70	2.63	5.47	5.34
		AAD of Snell normals (degree)	4.52	4.08	4.72	3.42	6.21	6.09
		RANSAC position inliers (%)	45.13	52.88	45.29	57.64	43.95	39.66
(c)		Mean positional error (mm)	0.17	0.21	0.15	0.12	0.13	0.21
		AAD of PCA normals (degree)	6.26	7.08	5.85	3.02	3.86	7.69
		AAD of Snell normals (degree)	6.86	6.76	6.43	3.88	4.82	7.22
		RANSAC position inliers (%)	36.29	41.62	37.37	50.10	44.52	38.58
(d)		Mean positional error (mm)	0.18	0.18	0.12	0.07	0.06	0.07
		AAD of PCA normals (degree)	6.36	5.37	3.31	1.90	1.71	1.75
		AAD of Snell normals (degree)	7.01	6.09	4.30	2.99	2.85	2.60
		RANSAC position inliers (%)	37.41	40.77	48.60	69.10	77.23	70.12

Table 1. Reconstruction errors of the “ornament” and “ball” objects. Several planar facets are manually labeled as shown in the images above. Each facet is fitted using RANSAC with the inlier threshold of 0.1mm . (a) and (b) show the results of the front and back surfaces of the ornament, respectively. (c) and (d) show the results of the front and back surfaces of the ball, respectively.

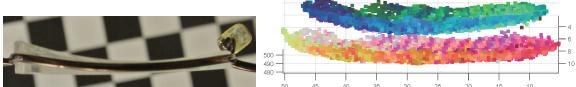


Figure 9. A failure case for a piece of myopia glass. The left image shows the top view of the object, whereas the right shows the reconstructed point cloud colored with PCA normals.

and not being able to handle colored objects. It is also noteworthy that the radiometric cues proposed in [7] may be incorporated to eliminate the monitor movements.

Essentially, our approach searches for a smooth surface that can best explain the observed ray-ray correspondences in terms of position-normal consistency. It implicitly assumes that there is only one feasible explanation for the observed correspondences. This assumption may not hold when the object is thin, in which case the refraction effects are mostly affected by the object thickness, rather than its shape. Hence, even though the reconstructed shape satisfies the position-normal consistency, it may not depict the real

object shape. Fig. 9 shows such a failure case.

Although only two cameras are used in our experiments, the proposed optimization procedure Eq.(3) can be extended to more than two views so that different cameras can fully cover the transparent objects. How to compute the PCA and Snell normals under such settings certainly deserves further investigation. It is noteworthy that Zuo *et al.* [42] recently develop a multi-view approach for reconstructing the full 3D shape of a transparent object, which requires numerous user interactions.

In addition to static transparent object reconstruction, we believe that the proposed position-normal consistency constraint is also applicable to dynamic wavefront reconstruction, where both the PCA and Snell normals are computable. We plan to apply the constraint to this problem in the near future as well.

Acknowledgments. We thank NSERC and the University of Alberta for the financial support. Constructive comments from anonymous reviewers and the area chair are highly ap-

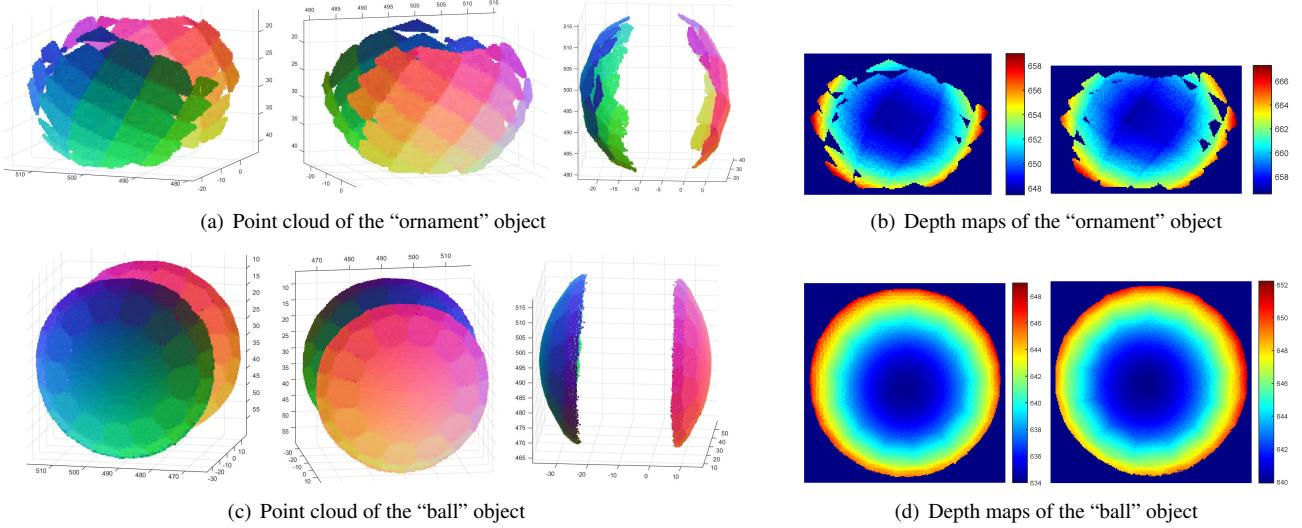


Figure 6. Reconstruction results of the “ornament” (top) and “ball” (bottom) objects; please refer to Table 1 for the photos. (a) and (c) show the 3D point clouds colored with the PCA normals seen from three different viewpoints. Both the front and the back points are plotted in the same coordinate. (b) and (d) show the depth maps of the front and the back surfaces. Note that some holes exist on the surface because no ray-ray correspondences are obtained for those regions.

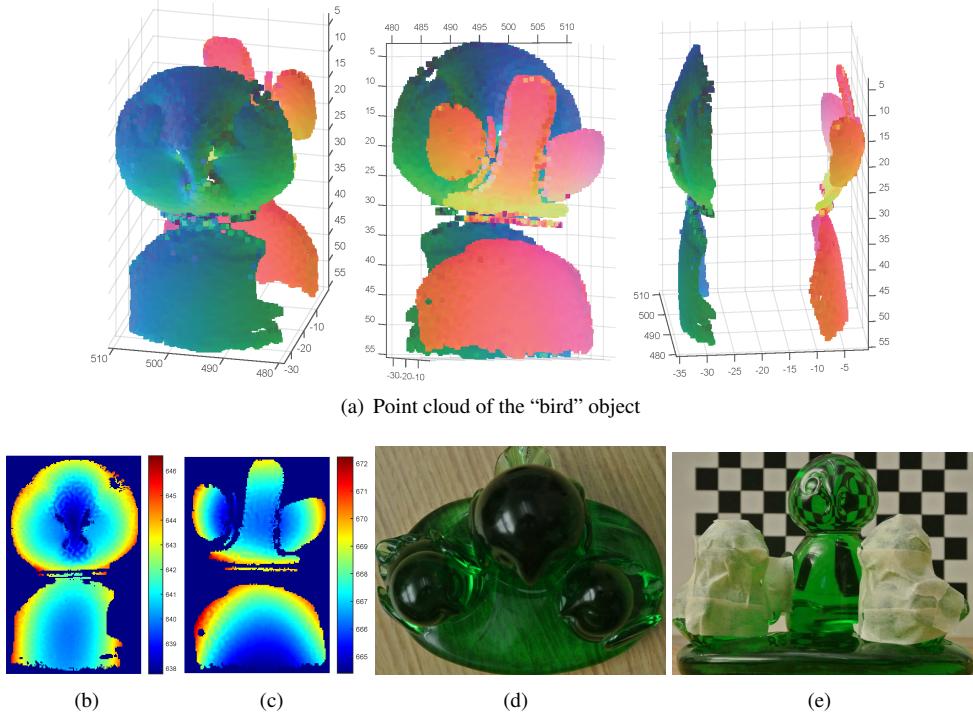


Figure 7. Reconstruction results of the “bird” object. (a) shows the point cloud colored with the PCA normals seen from three viewpoints. (b) and (c) show the depth map of the front and back surfaces. (d) shows the top view of the object. Because the positions of the three birds overlap, multiple refractions may happen between the camera and the light source. To avoid that, we cover the two smaller birds with tapes as shown in (e) and only reconstruct the larger bird for illustration. Note that the obtained back surface of the head of the larger bird is incomplete because the correspondences are not available in those complex regions.

preciated.

References

- [1] Project webpage. <https://webdocs.cs.ualberta.ca/~yqian3/papers/PNC/index.html>. 6

- [2] S. Agarwal, S. P. Mallick, D. Kriegman, and S. Belongie. On refractive optical flow. In *ECCV*, pages 483–494. 2004. 2
- [3] M. Alterman, Y. Swirski, and Y. Y. Schechner. Stella maris: Stellar marine refractive imaging sensor. In *ICCP*, 2014. 2
- [4] B. Atcheson, I. Ihrke, W. Heidrich, A. Tevs, D. Bradley, M. Magnor, and H.-P. Seidel. Time-resolved 3d capture of non-stationary gas flows. In *TOG*, page 132. ACM, 2008. 1
- [5] H. Averbuch-Elor, Y. Wang, Y. Qian, M. Gong, J. Kopf, H. Zhang, and D. Cohen-Or. Distilled collections from textual image queries. *Computer Graphics Forum*, 2015. 1
- [6] M. Ben-Ezra and S. K. Nayar. What does motion reveal about transparency? In *ICCV*. IEEE, 2003. 2
- [7] V. Chari and P. Sturm. A theory of refractive photo-light-path triangulation. In *CVPR*, pages 1438–1445, 2013. 7
- [8] Y.-Y. Chuang, D. E. Zongker, J. Hindorff, B. Curless, D. H. Salesin, and R. Szeliski. Environment matting extensions: Towards higher accuracy and real-time capture. In *SIGGRAPH*, pages 121–130, 2000. 2
- [9] Y. Ding, F. Li, Y. Ji, and J. Yu. Dynamic fluid surface acquisition using a camera array. In *ICCV*. IEEE, 2011. 1, 2
- [10] G. Eren, O. Aubreton, F. Meriaudeau, L. Sanchez Secades, D. Fofi, A. T. Naskali, F. Truchetet, and A. Ercil. Scanning from heating: 3d shape estimation of transparent objects from local surface heating. *Optics Express*, pages 11457–11468, 2009. 2
- [11] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. 5
- [12] M. Goesele, H. Lensch, J. Lang, C. Fuchs, and H.-P. Seidel. Disco: acquisition of translucent objects. In *TOG*, pages 835–844. ACM, 2004. 1
- [13] M. Gong, J. M. Selzer, C. Lei, and Y.-H. Yang. Real-time backward disparity-based rendering for dynamic scenes using programmable graphics hardware. In *Proceedings of Graphics Interface*, pages 241–248. ACM, 2007. 4
- [14] K. Han, K.-Y. K. Wong, and M. Liu. A fixed viewpoint approach for dense reconstruction of transparent objects. In *CVPR*, pages 4001–4008, 2015. 1, 5
- [15] M. B. Hullin, M. Fuchs, I. Ihrke, H.-P. Seidel, and H. P. Lensch. Fluorescent immersion range scanning. *TOG*, pages 87–87, 2008. 1
- [16] M. B. Hullin, M. Fuchs, I. Ihrke, H.-P. Seidel, and H. P. A. Lensch. Fluorescent immersion range scanning. In *SIGGRAPH*. ACM, 2008. 2
- [17] I. Ihrke, K. N. Kutulakos, H. Lensch, M. Magnor, and W. Heidrich. Transparent and specular object reconstruction. In *Computer Graphics Forum*, pages 2400–2426, 2010. 1, 2
- [18] I. Ihrke, K. N. Kutulakos, H. P. Lensch, M. Magnor, and W. Heidrich. State of the art in transparent and specular object reconstruction. In *EUROGRAPHICS*, 2008. 1, 2
- [19] Y. Ji, J. Ye, and J. Yu. Reconstructing gas flows using light-path approximation. In *CVPR*. IEEE, 2013. 1
- [20] K. N. Kutulakos and E. Steger. A theory of refractive and specular 3d shape by light-path triangulation. *IJCV*, pages 13–29, 2008. 1, 2, 3, 5, 6
- [21] D. Liu, X. Chen, and Y.-H. Yang. Frequency-based 3d reconstruction of transparent and specular objects. In *CVPR*, pages 660–667. IEEE, 2014. 1, 2
- [22] C. Ma, X. Lin, J. Suo, Q. Dai, and G. Wetzstein. Transparent object reconstruction via coded transport of intensity. In *CVPR*, pages 3238–3245. IEEE, 2014. 1
- [23] D. Miyazaki and K. Ikeuchi. Shape estimation of transparent objects by using inverse polarization ray tracing. *PAMI*, pages 2018–2030, 2007. 2
- [24] N. J. Morris and K. N. Kutulakos. Dynamic refraction stereo. In *ICCV*, pages 1573–1580. IEEE, 2005. 1, 2, 4
- [25] N. J. Morris and K. N. Kutulakos. Reconstructing the surface of inhomogeneous transparent scenes by scatter-trace photography. In *ICCV*, pages 1–8. IEEE, 2007. 1, 2
- [26] H. Murase. Surface shape reconstruction of an undulating transparent object. In *ICCV*, pages 313–317. IEEE, 1990. 2
- [27] Y. Qian, M. Gong, and Y. H. Yang. Frequency-based environment matting by compressive sensing. In *ICCV*. IEEE, 2015. 3, 5
- [28] R. B. Rusu. Semantic 3d object maps for everyday manipulation in human living environments. *KI-Künstliche Intelligenz*, 24(4):345–348, 2010. 3, 4
- [29] Q. Shan, S. Agarwal, and B. Curless. Refractive height fields from single and multiple images. In *CVPR*, pages 286–293. IEEE, 2012. 1, 2, 4
- [30] M. Tarini, H. P. Lensch, M. Goesele, and H.-P. Seidel. *3D acquisition of mirroring objects*. Max-Planck-Institut für Informatik, 2003. 2
- [31] B. Trifonov, D. Bradley, and W. Heidrich. Tomographic reconstruction of transparent objects. In *Proc. Eurographics Symposium on Rendering*, pages 51–60, 2006. 1, 2
- [32] C.-Y. Tsai, A. Veeraraghavan, and A. C. Sankaranarayanan. What does a light ray reveal about a transparent object? In *ICIP*. IEEE, 2015. 2, 6
- [33] G. Wetzstein, R. Raskar, and W. Heidrich. Hand-held schlieren photography with light field probes. In *ICCP*, pages 1–8. IEEE, 2011. 3, 6
- [34] G. Wetzstein, D. Roodnick, W. Heidrich, and R. Raskar. Refractive shape from light field distortion. In *ICCV*, pages 1180–1186. IEEE, 2011. 1, 2
- [35] Y. Wexler, A. W. Fitzgibbon, A. Zisserman, et al. Image-based environment matting. In *Rendering Techniques*, pages 279–290, 2002. 2
- [36] S. Wu, W. Sun, P. Long, H. Huang, D. Cohen-Or, M. Gong, O. Deussen, and B. Chen. Quality-driven poisson-guided autoscanning. *ACM Trans. Graph.*, 2014. 1
- [37] T. Xue, H. Mobahi, F. Durand, and W. T. Freeman. The aperture problem for refractive motion. In *CVPR*, 2015. 2
- [38] S.-K. Yeung, T.-P. Wu, C.-K. Tang, T. F. Chan, and S. Osher. Adequate reconstruction of transparent objects on a shoestring budget. In *CVPR*. IEEE, 2011. 2
- [39] Z. Zhang. A flexible new technique for camera calibration. *PAMI*, pages 1330–1334, 2000. 3
- [40] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal. Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software*, 23(4):550–560, 1997. 5
- [41] D. E. Zongker, D. M. Werner, B. Curless, and D. H. Salesin. Environment matting and compositing. In *SIGGRAPH*, pages 205–214, 1999. 2
- [42] X. Zuo, C. Du, S. Wang, J. Zheng, and R. Yang. Interactive visual hull refinement for specular and transparent object surface reconstruction. In *ICCV*, 2015. 7