This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the version available on IEEE Xplore.

Recovering Transparent Shape from Time-of-Flight Distortion

Kenichiro Tanaka[1,2] Yasuhiro Mukaigawa[2] Hiroyuki Kubo[2]
Yasuyuki Matsushita[1] Yasushi Yagi[1]
[1]Osaka University [2]Nara Institute of Science and Technology
{{tanaka,yagi}@am.sanken,yasumat@ist}.osaka-u.ac.jp {mukaigawa,hkubo}@is.naist.jp

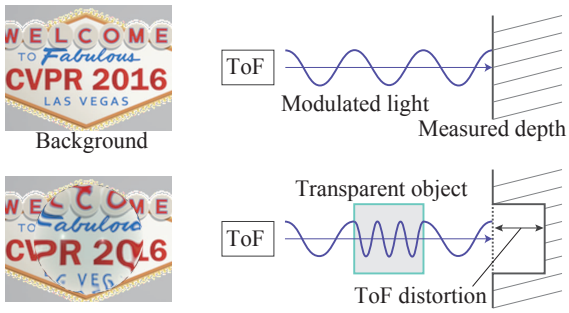
Abstract

This paper presents a method for recovering shape and normal of a transparent object from a single viewpoint using a Time-of-Flight (ToF) camera. Our method is built upon the fact that the speed of light varies with the refractive index of the medium and therefore the depth measurement of a transparent object with a ToF camera may be distorted. We show that, from this ToF distortion, the refractive light path can be uniquely determined by estimating a single parameter. We estimate this parameter by introducing a surface normal consistency between the one determined by a light path candidate and the other computed from the corresponding shape. The proposed method is evaluated by both simulation and real-world experiments and shows faithful transparent shape recovery.

1. Introduction

Transparent shape reconstruction is important for scientific imaging and applications in industrial manufacturing. It has been a difficult problem in computer vision because the appearance of a transparent object can only be indirectly observed by the distortion of background textures as illustrated in Fig. 1a. Several methods that uses the observations of the *geometric distortion* have been proposed, yet it is still an active research subject.

Recently, a Time-of-Flight (ToF) camera, which measures distance by correlation of its modulated light, is becoming a commodity device. When a scene containing a transparent object is recorded by a ToF camera, the ToF measurement also becomes distorted because the light slows down inside the transparent object due to its refractive index as illustrated in Fig. 1b. The distortion is different than the geometric distortion on the image coordinates, but still conveying rich information about the shape of transparent object. We call this distortion a *Time-of-Flight distortion*[1] in this paper and use it for recovering the shape of transparent object.



(a) Geometric distortion (b) Time-of-Flight distortion

Figure 1: Two distortions caused by transparent objects. (a) Background texture is geometrically distorted by refraction. (b) Measured depth is distorted by slower light speed in the medium.

Our method records two ToF measurements of a transparent object from a single viewpoint but by moving the background reference surface which is calibrated. Assuming that the refractive index of the target object is known and the ToF camera is calibrated, the shape estimation problem can be viewed as the problem of searching the light path, which correspond to estimating front and back refraction points and its surface normal. We show that, using the ToF distortion, the light path has a simple expression that is governed by a single parameter. We develop a method for estimating this parameter using a *surface normal consistency*, that represents a consistency between the surface normal computed from the light path candidate and that obtained from the corresponding shape.

The proposed method estimates both front and back surfaces in a single viewpoint approach. Unlike previous single viewpoint approaches that are restricted to a scene with a single refraction, or requiring a number of light sources to illuminate the scene, the proposed method is able to recover a scene with two refraction surfaces from a single view point, with two observations obtained by moving the

[1]This is not ordinary depth distortions that are due to calibration or multi-path effects but the distortion of "time-of-flight."

background reference surface. This new setting is enabled by the use of ToF distortion, which explicitly encodes the altered light speed in the transparent medium and its volume. Furthermore, we show a simple multi-path mitigation technique using a retroreflective sheet for this setting, which does not require any computational illumination devices, for recovering curved or multi-planar surfaces of transparent objects.

2. Related work

Early works of recovering the shape of transparent objects include [17], which recovers a single refraction water surface by observing the image placed under water. The setting of single refraction scenes has been further studied by several researchers. Morris and Kutulakos [16] reconstruct the shape of a dynamic wavy surface by observing a reference pattern placed under water from stereo cameras. Alterman *et al*. [1] estimate the position of the target object in the air from a camera placed in water using a stereo image sequence. Tian and Narasimhan [20] simultaneously estimate the shape of water surface and planar underwater scene from an image sequence by water surface tracking based on the distortion model using the wave equation. Tian and Narasimhan [21] also remove distortion of wavy surface in a data-driven approach and reconstruct wavy surface by spatially integrating the water distortion. Wetzstein *et al*. [22] reconstruct thin transparent objects by assuming thin transparent objects as a single refraction surface using light-field probe, which converts the position and angle of the light source into color codes. In contrast to these approaches that assume a single refraction path, our method focuses on scenes with two refraction paths to estimate whole shape.

There are also methods that analyze solid transparent shapes that exhibit two or more refractions. Kutulakos and Steger [12] show a general theory of tractability of shape recovery based on refractive paths characterized by the number of viewpoints, reference points, and refraction points. They also show reconstruction of a transparent shape from three viewpoints, two reference points, and two refraction points. Their problem setting is similar to ours but our method reconstructs from two observations from a single viewpoint using a ToF camera, with moving the background reference surface. There are also some single viewpoint approaches for transparent shape reconstruction. Morris and Kutulakos [15] reconstruct inhomogeneous transparent objects by illuminating an object from various positions and analyzing the specular reflections. Similarly, Yeung *et al*. [23] also propose a reconstruction method based on specular reflection analysis. While effective, these methods require a number of light positions for accurate recovery.

Some other methods take an approach of using specialized devices for recovering transparent shape. Han *et al*. [6] reconstruct a single side of transparent surface object by the crossing of two reference rays, one of which is measured in the air and the other is measured in liquid. There are also other unique methods for recovering transparent surfaces. Ihrke *et al*. [9] reconstruct the shape of flowing water by dyeing water with a fluorescent chemical and observing from multiple video cameras. Ma *et al*. [13] acquire the refractive index field based on the transport of intensity equations, which is a theory of phase imaging with coherent illumination, using collimated illumination, and reconstruct 3-d refractive volume based on tomography. Miyazaki and Ikeuchi [14] propose an inverse polarization ray-tracing for estimating the front surface of a transparent object using polarized reflections. Eren *et al*. [4] use thermal imaging for determining transparent shape by illuminating the target by laser beam.

	# view	# ref.	specialized device
[12]	3	2	–
[15, 23]	1	many	–
[22]	1	1 (+angle)	light-field probe
[6]	1	4	liquids
Ours	1	2	(retroreflective screen)

Table 1: Relation to the prior work. Our method recovers transparent shape from single viewpoint and two reference points without specialized devices.

Table 1 summarizes the settings of transparent shape estimation methods, mainly single viewpoint approaches. Our method determines the shape of transparent objects from a single viewpoint with two reference points. We consider that a ToF camera is now a commodity device because it is available at a similar cost with ordinary RGB cameras.

Similar to our setting, there are works that use a ToF camera with scenes including transparent objects. Heide *et al*. [7] recover light propagation sequences of a scene by sweeping the modulation frequency of their custom ToF camera. O'Toole *et al*. [18] separate light-in-flight images into direct reflection, specular inter-reflection, and global components such as caustics based on spatial probing. Kadambi *et al*. [11] reconstruct time sequential images of a scene including transparent objects by altering ToF measurement using coded light pulses. Gkioulekas *et al*. [5] visualize propagation of light by an optical coherence tomography scheme using a Michelson interferometer. Because their method has pico-second time resolution, spectral dispersion in a glass slab can be visualized. While these works visualize light transport of the scene, our goal is reconstruct the shape of transparent objects. Similar to our work, Shim and Lee [19] reconstruct the shape of thin *translucent* objects using a single ToF camera. They model

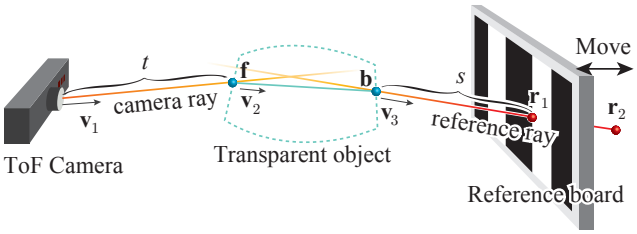


Figure 2: Transparent shape recovery problem. The front surface point $\mathbf{f}$ is on the camera ray at distance t, and the back surface point $\mathbf{b}$ is on the ray of 2 reference points at distance s.

the distortion of the depth by two layer (2-sparse) multi-path model, and determined object and background layer's depth from multiple observations of different illumination phase delay. Unlike their method, our method is developed for recovering shape of *transparent* objects, which have two refraction points and considerable thickness, using the variation of speed of light inside the object and refractive paths.

3. Proposed method

In our setting, we measure a transparent object using a known reference board placed behind the object as illustrated in Fig. 2. The target scene is recorded by a ToF camera twice by moving the reference board at two distinct locations behind the object. A single ToF observation contains both intensity and depth measurements, and from the intensity measurement, we determine two reference points $\mathbf{r}_1, \mathbf{r}_2 \in \mathbb{R}^3$. From these two reference point observations, the reference ray direction $\mathbf{v}_3$ can be determined. With an assumption that the refractive index ν of the transparent object is known and that the camera is calibrated therefore the camera ray direction $\mathbf{v}_1$ is known, our goal is to estimate the front surface point $\mathbf{f}$ and back surface point $\mathbf{b}$ at every camera pixel using the ToF depth measurement l_{ToF}. This problem is equivalent to estimating two variable t and s, where t is the distance from the camera to the front surface point, and s is the distance from the back surface point to the reference point. Our method estimates these unknowns t and s from $\mathbf{v}_1$, $\mathbf{v}_3$, $\mathbf{r}_1$, and l_{ToF} for determining the shape of transparent object. We begin with describing the ToF distortion model and develop the estimation method.

3.1. Time-of-Flight distortion model

A ToF sensor acquires the scene depth by observing the delay of returned light. When a transparent object is placed in the scene, a ToF sensor measures the optical length to the background object instead of the transparent object because light refracts and passes through the transparent object and reflects back from the background. In addition, the light speed slows down inside the transparent object according to its refractive index ν, hence the optical length measured by the ToF camera becomes different than the geometric length of the light path. The measured depth l_{ToF} can, therefore, be expressed as

$$l_{ToF} = t + \nu \left|\mathbf{b} - \mathbf{f}\right| + s, \tag{1}$$

where $|\cdot|$ represents the geometric length of the vector. We call this distortion the *Time-of-Flight distortion* in this paper, and use it for estimating the shape of transparent objects as it embeds the thickness of transparent object.

This model is built upon the following two assumptions.

- Reflection on the transparent object surface is ignored. Specular reflection is only observed when the surface confronts with the ToF camera, which rarely happens in practice.
- We assume only a single path of ToF light rays and ignore multi-path interference. In practice, we avoid the multi-path effect using a retroreflective sheet, which is explained in an experiment section. Prior methods for multi-path separation in ToF sensing such as [18, 11] can be alternatively used.

3.2. Baseline method

By examining the equation of the ToF distortion (Eq. (1)) analytically, s can be expressed as a function of t as[2]

$$s(t) = \frac{-h(t) - \sqrt{h^2(t) - gi(t)}}{g}, \tag{2}$$

where g, h, and i are auxiliary variables defined as

$$\begin{cases} g & = \nu^2 - 1 \\ h(t) & = l_{ToF} - t - \nu^2(\mathbf{r}_1 - t\mathbf{v}_1)^T\mathbf{v}_3 \\ i(t) & = \nu^2 \left|\mathbf{r}_1 - t\mathbf{v}_1\right|^2 - (l_{ToF} - t)^2. \end{cases} \tag{3}$$

While this expression eliminates unknown vectors $\mathbf{b}$ and $\mathbf{f}$, there still remains one degree of freedom to determine the unique shape s and t. For resolving the ambiguity, we use a surface normal consistency described in the following.

When a hypothesized depth t is assumed, we can obtain a hypothesized front surface point $\mathbf{f}$ and back surface point $\mathbf{b}$ as functions of t:

$$\begin{cases} \mathbf{f}(t) & = t\mathbf{v}_1, \\ \mathbf{b}(t) & = \mathbf{r}_1 - s(t)\mathbf{v}_3. \end{cases} \tag{4}$$

Based on this, the refractive ray direction $\mathbf{v}_2$ can be obtained from the hypothesized front surface point $\mathbf{f}$ to back surface point $\mathbf{b}$ as

$$\mathbf{v}_2(t) = \frac{\mathbf{b}(t) - \mathbf{f}(t)}{\left|\mathbf{b}(t) - \mathbf{f}(t)\right|}. \tag{5}$$

[2]Please refer to Appendix A in the supplementary material for the derivation.

Since the refractive index ν is known, the surface normal $\mathbf{n_p}(t)$ of the front surface point can be obtained from the refractive path using Snell's law as[3]

$$\begin{aligned}\mathbf{n_p}(t) &= \frac{\nu\mathbf{v}_2(t) - \mathbf{v}_1}{|\nu\mathbf{v}_2(t) - \mathbf{v}_1|} \\ &= \frac{1}{N_2(t)}\left(\frac{\nu}{N_1(t)}\left(\mathbf{r}_1 - s(t)\mathbf{v}_3 - t\mathbf{v}_1\right) - \mathbf{v}_1\right),\end{aligned} \tag{6}$$

where $N_1(t) = |\mathbf{b}(t) - \mathbf{f}(t)|$ and $N_2(t) = |\nu\mathbf{v}_2(t) - \mathbf{v}_1|$ are normalization coefficients.

At the same time, we can obtain another surface normal $\mathbf{n_d}(t)$ from the hypothesized shape as

$$\mathbf{n_d}(t) = \frac{\frac{\partial(t\mathbf{v}_1)}{\partial x} \times \frac{\partial(t\mathbf{v}_1)}{\partial y}}{\left|\frac{\partial(t\mathbf{v}_1)}{\partial x} \times \frac{\partial(t\mathbf{v}_1)}{\partial y}\right|}, \tag{7}$$

where $\times$ and ∂ are cross product and partial differentiation operators, respectively. We assume that the neighbors around $t\mathbf{v}_1$ correspond to those viewed in the camera pixel coordinates.

If the assumed front depth t is correct, two normals $\mathbf{n_p}(t)$ and $\mathbf{n_d}(t)$ should coincide; therefore, the estimation problem can be casted as an optimization problem as

$$\underset{\mathbf{t}}{\operatorname{argmin}} \sum_{c\in C} \|\mathbf{n}_{\mathbf{p},c}(t_c) - \mathbf{n}_{\mathbf{d},c}(t_c)\|_2^2, \tag{8}$$

where $\mathbf{t}$ is a vector listing t_c for all pixels, C is a set of all pixels, t_c is the hypothesized front depth of pixel c, $\mathbf{n}_{\mathbf{p},c}$ and $\mathbf{n}_{\mathbf{d},c}$ are the surface normal computed from the refractive path and that from hypothesized shape at pixel c, respectively. We call this method a *baseline method*, and in the next subsection, we introduce additional objectives for a more stable solution.

3.3. Robust estimation method

The objective function (Eq. (8)) directly uses the optical length measured by the ToF sensor (Eq. (1)) without considering uncertainty (or noise) in the measurement. To take into account observation noise, we introduce a new variable l_c for each pixel c, which represents the noise-free ToF optical length. There are two objective terms; one of which is the normal consistency used for the baseline method described above, and the other is that the denoised signal l_c should remain close enough to the measured signal l_{ToF}. We additionally regularize the denoising part assuming the smoothness of both front and back surfaces. Hence, the

[3]Please refer to Appendix B in the supplementary material.

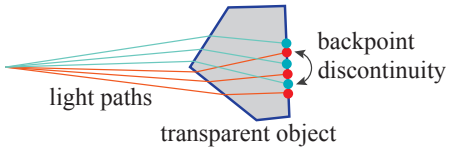


Figure 3: Discontinuity of back surface points. Red and blue points are well ordered in each group, but generates discontinuity due to the edge on the front surface as emphasized by a black arrow. To take into account of this discontinuity, we use a Huber cost function for back surface smoothness.

overall objective function becomes

$$\begin{aligned}\hat{\mathbf{t}}, \hat{\mathbf{l}} = \underset{\mathbf{t},\mathbf{l}}{\operatorname{argmin}} &\sum_{c\in C} \|\mathbf{n}_{\mathbf{p},c}(t_c, l_c) - \mathbf{n}_{\mathbf{d},c}(t_c)\|_2^2 + \\ &\lambda_1 \sum_{c\in C} \|l_c - l_{ToF}(c)\|_2^2 + \\ &\lambda_2 \sum_{j,k\in N} \|t_j\mathbf{v}_{1,j} - t_k\mathbf{v}_{1,k}\|_2^2 + \\ &\lambda_3 \sum_{c\in C} \left\|\frac{\partial}{\partial z}\mathbf{b}_j(t_c, l_c)\right\|_H,\end{aligned} \tag{9}$$

where $\mathbf{l}$ is a vector listing l_c for all pixels, j and k are pixel indices chosen from a set of all neighborhood N, $\|\cdot\|_H$ is the Huber penalty function [8][4], $\mathbf{v}_{1,c}$ is a unit camera ray vector corresponding to pixel c, and $l_{ToF}(c)$ is the measured ToF depth at pixel c. The first term ensures the normal consistency, the second term represents denoising of measured signal l_{ToF}, and the third and fourth terms regularize the denoising process by enforcing smoothness of front and back surfaces. If the front surface consists of multiple planes or is curved, continuity of back surface points could break as illustrated in Fig. 3. We therefore use a Huber cost function for the back surface smoothness because it allows occasional discontinuity (outliers) while retaining overall smoothness.

3.4. Solution method

The optimization problem of Eq. (9) is unfortunately non-convex and difficult to directly solve. We therefore take an alternating minimization approach by splitting the original problem into two subproblems; one for estimating $\mathbf{t}$ with keeping $\mathbf{l}$ fixed ($\mathbf{t}$-subproblem), and the other for determining $\mathbf{l}$ with given $\mathbf{t}$ ($\mathbf{l}$-subproblem). This can be interpreted as alternating estimation of shape ($\mathbf{t}$-subproblem)

[4] The Huber penalty function is an l_1/l_2 hybrid norm hence it can be used as smoothness with discontinuity, and defined as

$$\|\mathbf{x}\|_H = \sum_i h_\epsilon(x_i) \text{ , where } h_\epsilon(x) = \begin{cases} |x| - \epsilon/2 & (|x| > \epsilon) \\ x^2/(2\epsilon) & \text{otherwise} \end{cases}$$

and denoising of measurement (**l**-subproblem). Our method iteratively updates $\mathbf{t}$ and $\mathbf{l}$ with initialization of $\mathbf{t} = \texttt{constant}$, which is chosen manually by setting the approximate distance to the object, and $\mathbf{l} = \mathbf{l}_{ToF}$, which is a vectorized $l_{ToF}(c)$ for all pixels. We now discuss these two subproblems in detail.

t-subproblem By fixing $\mathbf{l}$, Eq. (9) can be reduced to an optimization problem of $\mathbf{t}$ as

$$\hat{\mathbf{t}} = \underset{\mathbf{t}}{\operatorname{argmin}} \sum_{c \in C} \left\| \mathbf{n}_{\mathbf{p},c}(t_c, \hat{l}_c) - \mathbf{n}_{\mathbf{d},c}(t_c) \right\|_2^2 + \lambda_2 \sum_{j,k \in N} \|t_j \mathbf{v}_{1,j} - t_k \mathbf{v}_{1,k}\|_2^2 , \tag{10}$$

where $\hat{l}_c$ represents the estimate of l_c obtained from the previous iteration. This problem can be interpreted as estimation of the front surface shape because t_c corresponds to the depth of front surface. Neglecting the back surface smoothness term is justified because of the fact that the length between $\mathbf{f}(t)$ and $\mathbf{b}(t)$ is almost fixed when $\hat{l}_c$ is unchanged. Due to the nested normalization terms as shown in Eq. (6), it is difficult to analytically derive its first and second-order derivatives of the objective function. To avoid high computational complexity of calculating second derivatives, we use the L-BFGS method [24] which only uses approximate Hessians rather than explicitly computing them. The $\mathbf{t}$-subproblem is again non-convex; therefore, there is a chance of being trapped by a local minima. As we will see in Sec. 4.1, in practice, it yields good estimates with appropriate initialization.

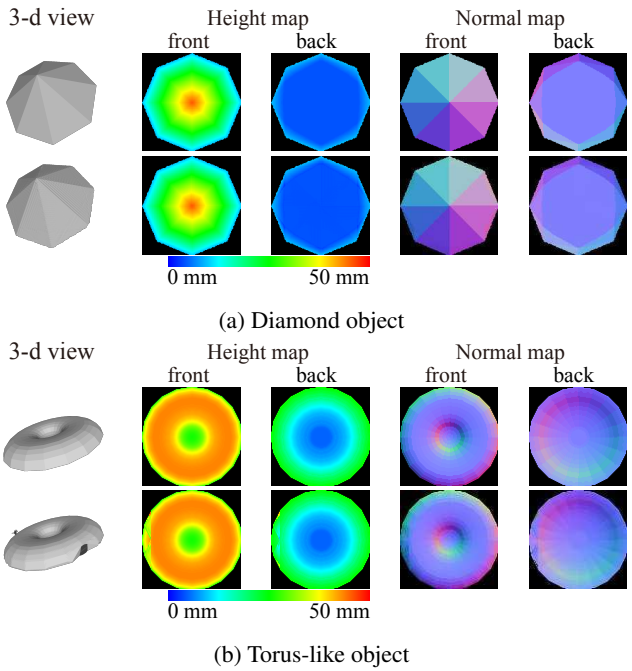
l-subproblem By fixing $\mathbf{t}$ and neglecting the front surface normal consistency, we obtain an $\mathbf{l}$ optimization problem written as

$$\hat{\mathbf{l}} = \underset{\mathbf{l}}{\operatorname{argmin}} \sum_{c \in C} \|l_c - l_{ToF}(c)\|_2^2 + \lambda_3' \sum_{c \in C} \left\| \frac{\partial}{\partial z} \mathbf{b}_j(\hat{t}_c, l_c) \right\|_H , \tag{11}$$

where $\lambda_3' = \lambda_3 / \lambda_1$. It can be viewed as a problem of denoising $\mathbf{l}$ regularized by the back surface smoothness (the second term) with a fixed front surface $\hat{\mathbf{t}}$. Although this problem is not strictly convex, we observed that this problem is approximately convex[5]; hence the optimal solution can be obtained. We again use the L-BFGS method as a minimizer for $\mathbf{l}$-subproblem.

The solution method for the baseline method described in Sec. 3.2 corresponds to the $\mathbf{t}$-subproblem. On the other hand, for the robust estimation method in Sec. 3.3, $\mathbf{t}$- and $\mathbf{l}$-subproblems are repeatedly solved in an alternating manner. In the implementation, we begin with the $\mathbf{t}$-subproblem and iterate until convergence. We stop the iteration when the gap between estimates and former estimates for both $\hat{\mathbf{t}}$ and $\hat{\mathbf{l}}$ are small enough.

[5]Please refer to Appendix C in the supplementary material for the convexity of $\mathbf{l}$-subproblem.



(a) Diamond object

(b) Torus-like object

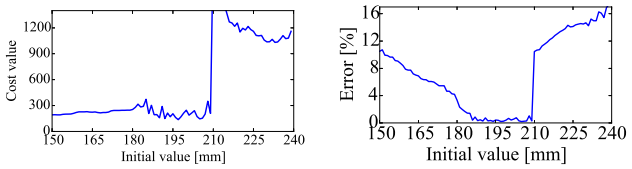
Figure 4: Simulation examples: Diamond and torus-like objects. The top row shows the ground truth of 3-d view, pseudo-colored height map, and normal map of both front and back surfaces. The bottom row shows the reconstruction result. The reconstruction error is 0.17% and 0.26%, respectively..

4. Experiments

In this section, we first assess the accuracy, robustness, and convergence of the proposed method using simulation data, and apply the method to real-world transparent objects to evaluate its effectiveness.

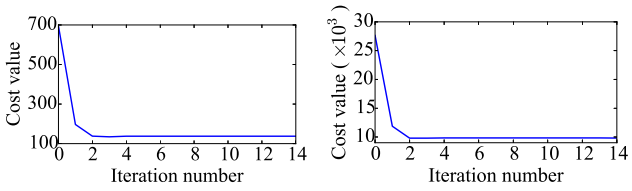
4.1. Simulation test

We generate simulation data of a scene with a transparent object that consist of the optical length and two reference points by ray tracing using Eq. (1) and Snell's law. To the data, we apply the proposed method to estimate front and back surface points and assess its accuracy by comparing with the ground-truth model. We use 0.005 for λ_2 and 20 for λ_3' for all target objects in the simulation tests.



(a) Cost value (b) Error of estimated shape

Figure 5: Assessment result about local minimum depending on the initial value. (a) Terminated cost value. (b) Error of estimated shape compared to the ground-truth. All initial values between $186\,\mathrm{mm}$ and $209\,\mathrm{mm}$ result in accurate shapes.



(a) Cost of t-subproblem (b) Cost of l-subproblem

Figure 6: Convergence of our alternating minimization. (a) Cost of t-subproblem. (b) Cost of l-subproblem. They decrease over iterations and converge after 3 iterations in this example.

Accuracy We first assess the effectiveness of the normal consistency objective (t-subproblem) using 48 types of transparent shapes generated by simulation. The data is noise-free in this experiment, and the solution is derived by solving the t-subproblem (baseline method with smoothness). The initial value of $\mathbf{t}$ is set to constant using an approximate depth to the object, which essentially corresponds to a planar surface located nearby the object. Figure 4a shows the result of a diamond shape scene. The result is close to the ground truth, whose root mean squared error (RMSE) is $0.524\,\mathrm{mm}$ (0.17 % of the optical length). As another example, Fig. 4b shows the result of a round torus-like object. The reconstruction accuracy is high in this case as well with RMSE $0.801\,\mathrm{mm}$ (0.26 % error), while there is visible artifacts at a few boundary regions. The average error of all 48 target objects is 0.45 %, and the result indicates the effectiveness of the normal consistency.

Effect of initialization We assess the influence of different initializations to the solution of t-subproblem because the t-subproblem is non-convex as described in Sec. 3.4. The transparent shape that we use for this test is placed at $200\,\mathrm{mm}$ from the camera, and its thickness is $50\,\mathrm{mm}$; thus, the object spans in the range of $200\,\mathrm{mm}$ and $250\,\mathrm{mm}$. We vary the initialization constant for $\mathbf{t}$ from $150\,\mathrm{mm}$ to $250\,\mathrm{mm}$ and assess the convergence and accuracy by solving the t-subproblem. Figure 5a shows the cost value at convergence, and Fig. 5b shows the reconstruction error. The reconstruction error is less than 1% in a wide range of initial values between $186\,\mathrm{mm}$ and $209\,\mathrm{mm}$, which shows the tolerance of the method against inaccurate initial values.

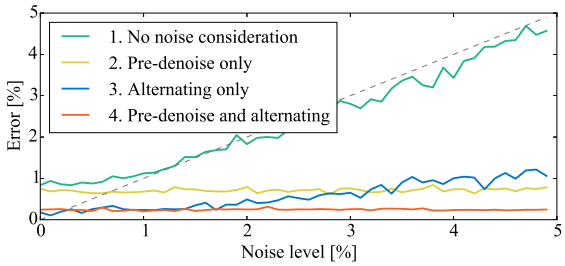


Figure 7: Reconstruction error of four approaches with respect to varying noise level. The reconstruction error generally becomes higher with a greater noise level. Errors are suppressed by proper noise handling. The combination of denoising and alternating optimization yields stable and the lowest error among these four approaches.

Convergence of alternating optimization We also assess the convergence of the alternating optimization described in 3.4. In this simulation, we add Gaussian noise with its standard deviation of 0.5% of optical length to all the pixels of the depth data. Figure 6 shows an example of the cost variations of two subproblems over iterations. The cost of both subproblems rapidly decreases at the beginning of iterations and remains stable after 3 iterations. With our test, most transparent shapes showed a similar convergence behavior and we consider that it is safe to say it reaches a local optimum.

Robustness against noise We further assess the effect of observation noise by adding noise to simulated ToF measurements. For comparison, we assess four approaches: (1) baseline method with smoothness (t-subproblem only), (2) t-subproblem with denoising, (3) alternating optimization (Sec. 3.4), and (4) alternating optimization with denoising. We use the Non-local Means Denoising [3] for denoising ToF measurements. Figure 7 summarizes the reconstruction errors of these four strategies with respect to the noise level. Without a proper noise handling (approach 1), results are significantly affected by observation noise hence the error increases together with the observation noise level. When denoising is applied prior to the optimization (approach 2), the error becomes stable regardless of the noise level. With the alternating optimization method (approach 3), the error is also suppressed while it shows degradation at higher noise levels. The highest accuracy is obtained by alternat-

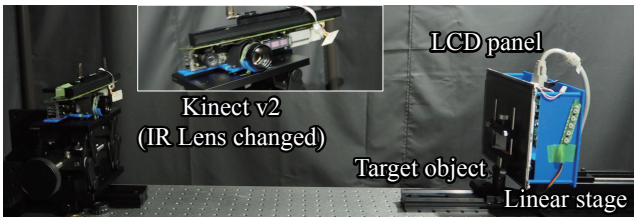


Figure 8: Experimental setup. IR lens of Kinect v2 is changed to obtain narrower field of view. Reference points are obtained using the LCD panel on a linear stage.

ing optimization with denoising (approach 4). The result indicates that denoising is effective, that alternating optimization is also effective in suppressing noise, and that the combination of these two is effective. For the real-world experiments in the next subsection, we use the alternating optimization with denoising approach.

4.2. Real-world experiment

For the real-world experiment, we use an off-the-shelf ToF camera (Kinect v2) and an LCD panel placed on a linear stage. Three distinct transparent objects are used for conducting the experiment. The parameters are set to $\lambda_2 = 0.05$ and $\lambda_3' = 10$ for all target objects in this experiment.

Setup Figure 8 shows our experimental setup. We use Microsoft Kinect v2 for a ToF camera, whose lens is changed to Edmund 35mm IR lens to narrower the field of view (FOV). To obtain reference points, we use an LCD panel, which is mounted on a motorized linear stage (OptoSigma SGSP26-150) that allows replication of positions in high precision. The LCD panel reflects the lights from Kinect with displayed patterns hence the back illumination of the display is turned off.

Calibration and measurements The LCD-camera system is calibrated at two LCD locations before measurement. The 3-d position of every LCD's pixels at two depth locations are measured in the form of IR and depth images of calibrated Kinect. The pixel location of the LCD panel is determined by Gray code pattern projection method [10]. For measurement, we place the target transparent object in between the camera and LCD panel. The target object is measured twice with distinct LCD panel locations. To obtain reference points in the measurement, we again use the Gray code pattern projection.

Result We conduct experiments using three transparent objects, which do not produce multi-path interferences. Figure 9 shows target objects and the reconstruction results.

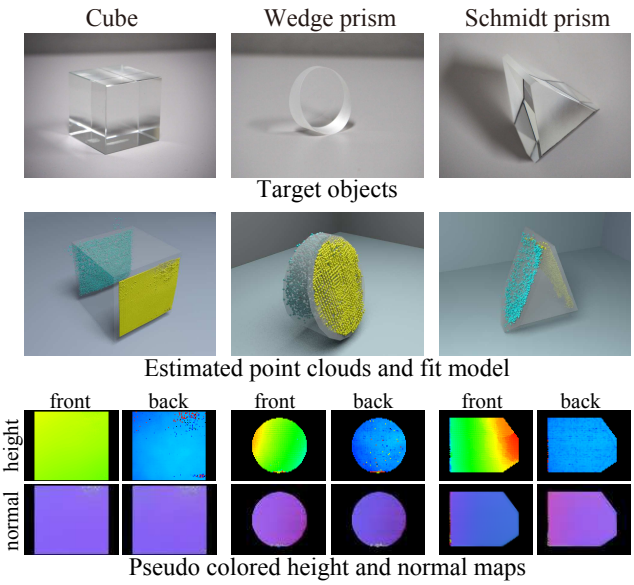


Figure 9: Experimental result of simple objects that consist of two planer surfaces. *Top:* Target objects. Front and back surfaces are single plane, hence no multi-path interferences are occurred. *Middle:* Estimated front points (yellow) and back points (cyan) with fit model by ICP (semi-transparent). *Bottom:* Estimated height map (upper) and normal map (lower) of front surface (left) and back surface (right).

Object	Mean	Std. dev.
Cube (parallel surfaces)	0.188 mm	0.458 mm
Wedge prism (18.8°)	0.226 mm	1.137 mm
Schmidt prism (45°)	0.381 mm	1.398 mm

Table 2: Quantitative evaluation of reconstruction errors. We evaluate the mean and standard deviation of the Euclidean distance between recovered points and the ground truth CAD model. The estimated points are registered to the model by ICP algorithm prior to the evaluation.

We can see that estimated point clouds well fit the ground-truth 3-d CAD model. To align the ground truth with the reconstruction, we use the Iterative Closest Point (ICP) [2] algorithm. The quantitative reconstruction errors are summarized in Table 2. The mean error in the Euclidean distance is small, and it shows the effectiveness of our method.

Multi-path avoidance If the target object is curved or has multiple planes, caustics appear on the background surface. In such a case, there are multiple light paths sharing the same reflection point on the background; therefore the observed depth by a ToF camera does not satisfy the model of Eq. (1). To avoid this multi-path interference, we use a

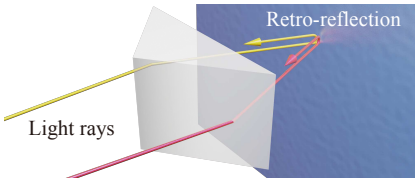
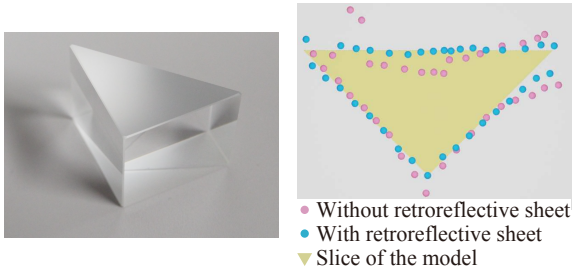


Figure 10: Multi-path avoidance using a retroreflective sheet. Using a retroreflective sheet, two rays (colored yellow and red) can be observed separately even when their reflection point is overlapped.



(a) Target object (b) Slice view

Figure 11: Reconstruction result of a right angle prism. (a) The target object. (b) a slice view of the estimated surface points and the ground truth. Blue and red points are the result with and without retroreflective sheet, respectively. The retroreflective sheet mitigates multi-path interference for more faithful recovery.

Object	Mean	Std. dev.
without retroreflection	0.745 mm	1.115 mm
with retroreflection	0.448 mm	0.828 mm

Table 3: Numerical evaluation of right angle prism. The result with retroreflective sheet is more accurate than that without it.

retroreflective sheet placed on the background. The retroreflective sheet reflects the incident light ray to the incident direction; hence, multi-path interference can be avoided even when the reference points are overlapped as shown in Fig. 10.

To verify the effectiveness of the multi-path avoidance, we use a right angle prism shown in Fig. 11a measure it with and without a retroreflective sheet. Figure 11b shows a slice of the estimated points and the ground truth model. The result without a retroreflective sheet is distorted by multi-path interference while the result with the retroreflective sheet shows faithful recovery of the object. It is also verified quantitatively by the reconstruction errors of the experiment summarized in Table 3.

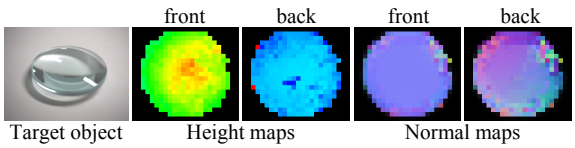


Figure 12: Reconstruction of a transparent curved object (convex lens)

Finally, we recover the shape of a convex lens, which has curved surfaces. We again use a retroreflective sheet for avoiding the multi-path effect. Figure 12 shows the estimated results. While it became somewhat noisy, the curved surface that exhibits multi-path refraction rays is recovered.

5. Discussions

We developed a method for transparent shape recovery using the time-of-flight distortion. One of the issues of our method for a practical use is that it is currently limited to low-resolution, because it is bounded by the resolution of the ToF image sensor. Another issue is that our method breaks down when the light path refracts more than twice, *e.g.*, due to total reflections inside the transparent object, which could occur at the edge of the object. To avoid this problem, the region near such edges should be treated differently by developing a suitable technique.

The normal consistency objective appeared in (8) and (9) is defined as the Euclidean distance of unit normal vectors. Theoretically, it should be better written by

$$\underset{\mathbf{t}}{\operatorname{argmin}} \sum_{c \in C} \left(\mathbf{n}_{\mathbf{p},c}(t_c)^T \mathbf{n}_{\mathbf{d},c}(t_c) - 1 \right)^2 ,$$

because of its directional nature. We have tested the above expression; however, the result did not change much while the computational cost increased significantly. Therefore, we decided to keep using the Euclidean distance of normals.

While we assume the neighbors of front surface points correspond to those viewed in the camera pixel coordinate, we cannot use the same assumption for the back surface because of refractions (depicted in Fig. 3), and a less restrictive smoothness term is defined for the back surface using a Huber cost function. We consider this design tends to make the front surface smoother than the back surface. This behavior can be balanced by adjusting weight parameters λs. Also, our solution is somehow heuristic approach for solving non-convex problem hence it is not guaranteed to reach a good estimate for any shapes.

In the future work, we are interested in exploring the direction of combining a computational illumination approach, such as [11], for not only mitigating the multi-path interference but also explicitly handling the multi-path effect for transparent shape recovery.

References

[1] M. Alterman, Y. Y. Schechner, and Y. Swirski. Triangulation in Random Refractive Distortions. In *Proc. International Conference on Computational Photography*, pages 1–10. IEEE, 2013. 2

[2] P. Besl and N. D. McKay. A Method for Registration of 3-D Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 14(2):239–256, 1992. 7

[3] A. Buades, B. Coll, and J.-M. Morel. A Non-Local Algorithm for Image Denoising. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 60–65. IEEE, 2005. 6

[4] G. Eren, O. Aubreton, F. Meriaudeau, L. A. 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*, 17(14):11457–11468, July 2009. 2

[5] I. Gkioulekas, A. Levin, F. Durand, and T. Zickler. Micron-scale Light Transport Decomposition Using Interferometry. *ACM Transactions on Graphics (ToG)*, 34(4), 2015. 2

[6] K. Han, K.-Y. K. Wong, and M. Liu. A Fixed Viewpoint Approach for Dense Reconstruction of Transparent Objects. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4001–4008. IEEE, 2015. 2

[7] F. Heide, M. B. Hullin, J. Gregson, and W. Heidrich. Low-budget Transient Imaging using Photonic Mixer Devices. *ACM Transactions on Graphics (ToG)*, 32(4), 2013. 2

[8] P. J. Huber. Robust Regression: Asymptotics, Conjectures and Monte Carlo. *The Annals of Statistics*, 1(5):799–821, Sept. 1973. 4

[9] I. Ihrke, B. Goidluecke, and M. Magnor. Reconstructing the Geometry of Flowing Water. In *Proc. International Conference on Computer Vision (ICCV)*, volume 2, pages 1055–1060. IEEE, 2005. 2

[10] S. Inokuchi, K. Sato, and F. Matsuda. Range Imaging System for 3-d Object Recognition. In *Proc. International Conference on Pattern Recognition*, pages 806–808, 1984. 7

[11] A. Kadambi, R. Whyte, A. Bhandari, L. Streeter, C. Barsi, A. Dorrington, and R. Raskar. Coded Time of Flight Cameras: Sparse Deconvolution to Address Multipath Interference and Recover Time Profiles. *ACM Transactions on Graphics (ToG)*, 32(6):1–10, Nov. 2013. 2, 3, 8

[12] K. Kutulakos and E. Steger. A Theory of Refractive and Specular 3D Shape by Light-path Triangulation. In *Proc. International Conference on Computer Vision (ICCV)*, volume 2, pages 1448–1455. IEEE, 2005. 2

[13] C. Ma, X. Lin, J. Suo, Q. Dai, and G. Wetzstein. Transparent Object Reconstruction via Coded Transport of Intensity. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3238–3245. IEEE, 2014. 2

[14] D. Miyazaki and K. Ikeuchi. Inverse Polarization Raytracing: Estimating Surface Shapes of Transparent Objects. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 910–917. IEEE, 2005. 2

[15] N. J. W. Morris and K. N. Kutulakos. Reconstructing the Surface of Inhomogeneous Transparent Scenes by Scatter-Trace Photography. In *Proc. International Conference on Computer Vision (ICCV)*, pages 1–8. IEEE, 2007. 2

[16] N. J. W. Morris and K. N. Kutulakos. Dynamic Refraction Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 33(8):1518–1531, Jan. 2011. 2

[17] H. Murase. Surface Shape Reconstruction of an Undulating Transparent Object. In *Proc. International Conference on Computer Vision (ICCV)*, pages 313–317. IEEE, 1990. 2

[18] M. O'Toole, F. Heide, L. Xiao, M. B. Hullin, W. Heidrich, and K. N. Kutulakos. Temporal Frequency Probing for 5D Transient Analysis of Global Light Transport. *ACM Transactions on Graphics (ToG)*, 33(4):1–11, July 2014. 2, 3

[19] H. Shim and S. Lee. Recovering Translucent Object using a Single Time-of-Flight Depth Camera. *IEEE Transactions on Circuits and Systems for Video Technology*, (99):1–14, 2015. 2

[20] Y. Tian and S. G. Narasimhan. Seeing through Water: Image Restoration using Model-based Tracking. In *Proc. International Conference on Computer Vision (ICCV)*, pages 2303–2310. IEEE, 2009. 2

[21] Y. Tian and S. G. Narasimhan. A Globally Optimal Data-driven Approach for Image Distortion Estimation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1277–1284. IEEE, 2010. 2

[22] G. Wetzstein, D. Roodnick, W. Heidrich, and R. Raskar. Refractive Shape from Light Field Distortion. In *Proc. International Conference on Computer Vision (ICCV)*, pages 1180–1186. IEEE, 2011. 2

[23] 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 *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2513–2520. IEEE, 2011. 2

[24] C. Zhu, R. H. Byrd, and J. Nocedal. L-BFGS-B: Algorithm 778: L-BFGS-B, FORTRAN Routines for Large Scale Bound Constrained Optimization. *ACM Transactions on Mathematical Software*, 23(4):550–560, 1997. 5