

Dynamic Refraction Stereo

Nigel J. W. Morris

Kiriakos N. Kutulakos

University of Toronto

Abstract

In this paper we consider the problem of reconstructing the 3D position and surface normal of points on an unknown, arbitrarily-shaped refractive surface. We show that two viewpoints are sufficient to solve this problem in the general case, even if the refractive index is unknown. The key requirements are (1) knowledge of a function that maps each point on the two image planes to a known 3D point that refracts to it, and (2) light is refracted only once. We apply this result to the problem of reconstructing the time-varying surface of a liquid from patterns placed below it. To do this, we introduce a novel “stereo matching” criterion called *refractive disparity*, appropriate for refractive scenes, and develop an optimization-based algorithm for individually reconstructing the position and normal of each point projecting to a pixel in the input views. Results on reconstructing a variety of complex, deforming liquid surfaces suggest that our technique can yield detailed reconstructions that capture the dynamic behavior of free-flowing liquids.

1. Introduction

Modeling the time-varying surface of a liquid has attracted the attention of many research fields, from computer graphics [1, 2] and fluid mechanics [3] to oceanography [4, 5]. While great strides have been achieved in the development of computer simulators that are physically accurate and visually correct [1], capturing the time-varying behavior of a real liquid remains a challenging problem.

From the point of view of computer vision, analyzing the behavior of liquids from videos poses several difficulties compared to traditional “3D photography” applications:

- *No prior scene model*: Spatio-temporal evolution is constrained only by the laws of fluid mechanics, making it difficult to assume a low-degree-of-freedom parametric model for such a scene [6].
- *Non-linear light path*: Liquid surfaces bend the incident light and, hence, a point below the surface will project along a non-linear path to a viewpoint above it.
- *Shape-dependent appearance modulation*: Absorption, scattering and Fresnel transmission cause the appearance of points below the surface to depend on the light’s path and, hence, on the surface shape [7].
- *Turbulent behavior*: Liquid flow is an inherently volumetric phenomenon whose complete characterization

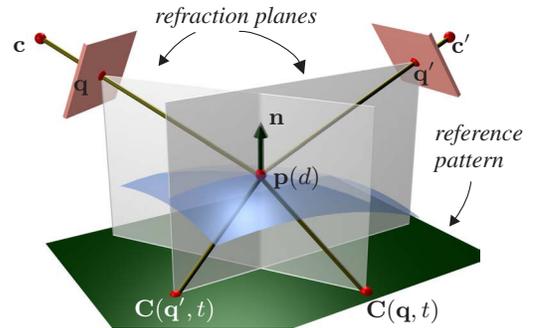


Figure 1. Geometry of refraction stereo. The goal is to reconstruct for each pixel \mathbf{q} , the 3D position and surface normal of point \mathbf{p} on the refractive surface.

requires capturing both its time-varying surface and a vector field describing internal motion [8].

- *Instantaneous 3D capture*: Since liquids are dynamic and can flow rapidly, shape recovery must rely on instantaneously-captured information.

As a first step, in this paper we consider the problem of reconstructing the time-varying 3D surface of an unknown liquid by exploiting its refractive properties. To do this, we place a known, textured pattern below the liquid’s surface and capture image sequences of the pattern from two known viewpoints above the liquid (Figure 1). Our focus is on imposing as few restrictions as possible on the scene—we assume that the liquid has a constant but unknown index of refraction and that its instantaneous 3D shape is arbitrary, as long as light coming from the pattern is refracted at most once before reaching the input viewpoints.

The reconstruction of refractive surfaces from photographs has a long history in photogrammetry [9–12]. These techniques assume a low-parameter model for the surface (e.g., a plane) and solve a generalized structure from motion problem in which camera parameters, surface parameters, and 3D coordinates of feature points below the surface are estimated simultaneously. In computer vision, the reconstruction of time-varying refractive surfaces was first studied by Murase [13], whose seminal work focused on water (whose refraction index is known) and followed a “shape-from-distortion” approach [4, 14]. In this approach, instantaneous 3D shape is recovered by analyzing one distorted image of a pattern that is placed underwater. Unfortunately it is impossible, in general, to reconstruct the 3D

shape of a general refractive surface from one image, even if its refractive index is known. The inherently ill-posed nature of the problem has prompted a variety of assumptions, including statistical assumptions about the pattern’s appearance over time [13], known average water height [4, 12], surface integrability [15], and special optics [5, 16]. These assumptions break down when the refractive index is unknown or when the liquid undergoes significant deformations that cause changes in both shape and height (e.g., pouring water in an empty tank).

A closely related problem is the reconstruction of highly specular surfaces, such as mirrors [15, 17–20]. Mirrors interact with light in much the same way that refractive surfaces do—light incident at a point is reflected according to the point’s surface normal, thereby tracing a non-linear path. Blake [17] proposed using a moving observer to recover the differential properties of a smooth mirror surface from the observed motion of specularities. Sanderson *et al* [21] were the first to analyze the ambiguities in single-view mirror reconstruction and to propose a stereo camera configuration for resolving them. Our work, which is based on a novel analysis of two-view ambiguities for refractive scenes, exploits some of the same basic insights. Recently, Bonfort and Sturm [22] used multiple views to build coarse volumetric models of mirror surfaces. While their work is promising, the development of algorithms able to recover detailed models of complex, unknown, mirror surfaces from multiple views remains open.

Reconstructing transparent liquid surfaces is even more challenging than mirrors for three reasons. First, the interaction between light and a mirror does not depend on the mirror’s material properties but it does depend on a liquid’s refractive index. When this index is unknown, it must be estimated along with 3D shape. Second, the non-linearity of light paths cannot be taken for granted in the case of fluctuating liquid surfaces, whose distance from a pattern below the surface may approach zero, diminishing the effect of refraction. To guarantee stable shape solutions, a reconstruction algorithm must be immune to such degeneracies. Third, establishing accurate pixel-wise correspondences between patterns and their distorted images is much easier in the case of a mirror. In liquids, the distortions are both geometric and radiometric (due to absorption, Fresnel effect, etc.) and can vary significantly from one instant to the next.

The starting point for our work is a novel geometrical result showing that two viewpoints are sufficient to compute both the shape and the refractive index of an unknown, generic refractive surface. The only requirements are (1) knowledge of a function that maps each point on the image plane to a known 3D point that refracts to it, and (2) light is refracted only once. Compared to mirrors, this is a stronger two-view result because it shows that the refractive index ambiguity, not present in mirror scenes, can be resolved without additional views.

On the practical side, our interest is in algorithms that can capture the detailed dynamic behavior of free-flowing liquids. To this end, our work has four contributions. First, we formulate a novel optimization criterion, called *refrac-*

tive disparity, appropriate for refractive scenes, that is designed to remain stable when refraction diminishes. Second, we develop an optimization-based algorithm for individually reconstructing the position and normal of each point projecting to the input views. The algorithm is closer to traditional triangulation [23] and bundle adjustment [24] than to voxel-based stereo [22], and imposes no constraints on the liquid’s shape or its evolution. Third, we show that refraction stereo can produce a detailed, full-resolution depth map and a separate, full-resolution normal map for the unknown surface. To our knowledge, only one other shape recovery method, Helmholtz stereopsis [25], has demonstrated the ability to compute dense normal maps along with some depth information (although its depth maps were low resolution and deemed inaccurate). Fourth, we present experimental results for a variety of complex, deforming liquid surfaces. These results suggest that refraction stereo can yield detailed reconstructions that capture the complexity and dynamic behavior of liquids.

2. Refraction Stereo Geometry

Consider an unknown, smooth, transparent surface that is viewed by two calibrated cameras under perspective projection (Figure 1). We assume that the surface bounds a homogeneous transparent medium (e.g., water or alcohol) with an unknown refractive index. Our goal is to compute the refractive index of the medium and the 3D coordinates and surface normal at each point on the unknown surface. To do this, we place a known reference pattern below the surface and compute a pixel-to-pattern correspondence function, $\mathbf{C}(\mathbf{q}, t)$, that gives us the 3D coordinates of the point on the pattern that refracts to pixel \mathbf{q} at time t . In the following, we assume that this function is known and concentrate on the instantaneous reconstruction problem at time t . We consider the problem of estimating the correspondence function in Section 5. To simplify notation, we omit the time parameter in the following discussion.

Let \mathbf{q} be a pixel in the input views, let $\mathbf{C}(\mathbf{q})$ be the point refracting to \mathbf{q} , and suppose that this refraction occurs at distance d from the image plane, at a point $\mathbf{p}(d)$ on the ray through pixel \mathbf{q} (Figure 1). The relation between pixel \mathbf{q} and points $\mathbf{C}(\mathbf{q})$ and $\mathbf{p}(d)$ is governed by Snell’s law, which describes how light is redirected at the boundary between two different media [7]. Snell’s law can be expressed as two independent geometric constraints:

- a *deflection constraint*, establishing a sinusoidal relation between incoming and outgoing light directions:

$$\sin \theta_o = r \sin \theta_i \quad (1)$$

where θ_i is the angle between the surface normal and the ray through $\mathbf{C}(\mathbf{q})$ and $\mathbf{p}(d)$; θ_o is the angle between the surface normal and the ray through pixel \mathbf{q} ; and r is the refractive index;

- and a *planarity constraint*, forcing the surface normal at $\mathbf{p}(d)$ to lie on the plane defined by point $\mathbf{C}(\mathbf{q})$ and the ray through \mathbf{q} ; we call this plane the *refraction plane* of pixel \mathbf{q} .

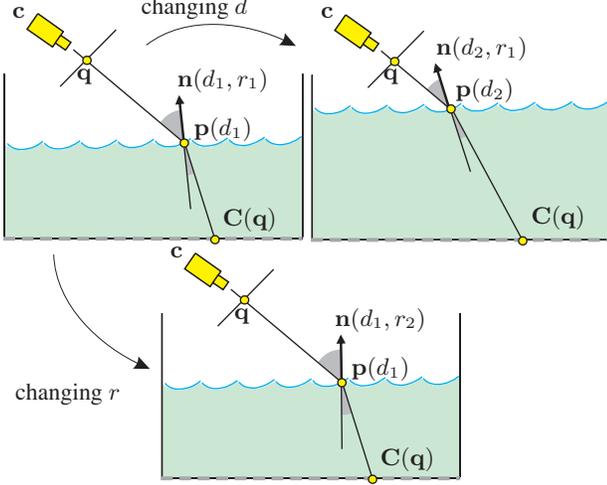


Figure 2. Single-viewpoint ambiguities. A view of pixel \mathbf{q} 's refraction plane is shown. Given a refractive index r_1 , we can find, for each distance to the surface, a normal that refracts point $\mathbf{C}(\mathbf{q})$ to its corresponding pixel \mathbf{q} . Similarly, given a distance d_1 , we can find, for each refractive index, a normal that refracts point $\mathbf{C}(\mathbf{q})$ to pixel \mathbf{q} .

These two constraints give us a relation between the pixel, a known 3D point that refracts to it, and the unknown surface. Unfortunately, they are not sufficient to determine how far from the image plane the refraction occurs, even when we do know the refractive index. This is because for every hypothetical distance there is a 1D set of possible normals that satisfy the planarity and deflection constraints. Each of these normals lies on the pixel's refraction plane and satisfies Eq. (1) for some value of the refractive index (Figure 2). Hence, the surface normals that satisfy Snell's law for pixel \mathbf{q} can be expressed as a two-parameter function, $\mathbf{n}(d, r)$, parameterized by the distance d and the unknown refractive index, r . A closed-form expression for this normal can be derived as

$$\mathbf{n}(d, r) = r \|\mathbf{i}(d) \wedge \mathbf{o}\| \left(\frac{\mathbf{i}(d) - [\mathbf{i}(d) \cdot \mathbf{o}] \mathbf{o}}{\|\mathbf{i}(d) - [\mathbf{i}(d) \cdot \mathbf{o}] \mathbf{o}\|} \right) + \left(r [\mathbf{i}(d) \cdot \mathbf{o}] - 1 \right) \mathbf{o} \quad (2)$$

where \wedge denotes vector product; \mathbf{o} is the direction of the ray through pixel \mathbf{q} ; and $\mathbf{i}(d)$ is the direction of the ray incident to the surface point $\mathbf{p}(d)$:

$$\mathbf{o} = \frac{\mathbf{c} - \mathbf{q}}{\|\mathbf{c} - \mathbf{q}\|}, \quad \mathbf{i}(d) = \frac{\mathbf{c} - d\mathbf{o} - \mathbf{C}(\mathbf{q})}{\|\mathbf{c} - d\mathbf{o} - \mathbf{C}(\mathbf{q})\|}. \quad (3)$$

When the refractive index has a known value r_0 , there is only one consistent normal, $\mathbf{n}(d, r_0)$, for each distance d . Sanderson *et al* [21] were the first to point out that this distance-normal ambiguity for a pixel \mathbf{q} can be resolved with the help of a second viewpoint.¹ Intuitively, a second

¹Sanderson *et al* [21] made this observation in the context of reconstructing opaque specular, rather than refractive, surfaces. Their analysis, however, applies equally well to the case of refractive surfaces with a known refractive index.

viewpoint allows us to “verify” whether or not a particular distance hypothesis d is correct (Figure 1): given such a hypothesis and given the projection \mathbf{q}' of point $\mathbf{p}(d)$ in the second camera, we simply need to verify that point $\mathbf{C}(\mathbf{q}')$ on the reference pattern refracts to pixel \mathbf{q}' .

While this hypothesis-verification procedure leads directly to an algorithm when the surface has a known refractive index, it leaves open the question of how to reconstruct surfaces whose 3D shape *and* refractive index are unknown. In this case, the surface normal lies in the full, two-parameter family, $\mathcal{N} = \{\mathbf{n}(d, r) \mid d, r \in \mathbb{R}^+\}$. One approach would be to use a third viewpoint to verify that a hypothetical refractive index r and distance d are consistent with the pixel-pattern correspondences in the three views.

Rather than use a third viewpoint, we prove that two views are, in fact, sufficient to estimate the 3D shape and refractive index of an unknown, generic surface. Intuitively, generic surfaces embody the notion of *non-degeneracy*—they are smooth surfaces whose differential properties remain unchanged if we deform their surface by an infinitesimal amount [26]. As such, they are especially suitable for modeling the complex, unconstrained shape of a liquid. Theorem 1 tells us if the liquid's surface is generic, the family, \mathcal{N} , of ambiguous solutions is discrete:

Theorem 1. \mathcal{N} is a zero-dimensional manifold for almost all pixels in the projection of a generic surface.

Theorem 1 suggests that it might be possible to compute the refractive index of a surface by choosing a single pixel \mathbf{q} and finding the distance and refractive index that are consistent with $\mathbf{C}(\mathbf{q})$ and the pixel-to-pattern correspondences in the second viewpoint. In practice, image noise and the possibility of multiple discrete solutions dictate an alternative strategy, where measurements from multiple pixels contribute to the estimation of the refractive index. We consider the algorithmic implications of this result below.

3. Dynamic Refraction Stereo Algorithm

In order to reconstruct a liquid's surface at a time instant t , we need to answer three basic questions: (1) how do we compute the pixel-to-pattern correspondence function, $\mathbf{C}(\mathbf{q}, t)$, (2) how do we compute the refractive index and (3) how do we assign a distance and a normal to each pixel?

To compute $\mathbf{C}(\mathbf{q}, t)$ we rely on a procedure that computes the correspondences for time $t = 0$ and then propagates them through time using optical flow estimation.

Since the refractive index is the same for all pixels, we seek a value that most closely satisfies the refractive stereo geometry across all pixels and all frames. To do this, we perform a discrete 1D search in an interval of plausible refractive indices and, for each hypothetical value, attempt to reconstruct the scene for all pixels and frames. We then choose the value that produces the smallest reconstruction error. This leads to the following general algorithm, whose steps are discussed in the following sections.

Step 1 Initialise pixel-to-pattern correspondences, $C(\mathbf{q}, 0)$.

Step 2 For each frame $t > 0$, estimate 2D optical flow to compute $C(\mathbf{q}, t)$ from $C(\mathbf{q}, t - 1)$.

Step 3 For every refractive index $r \in \{r_1, \dots, r_n\}$, every frame t and every pixel \mathbf{q} ,

- assuming refractive index r for the liquid, estimate the 3D position \mathbf{p} and normal \mathbf{n} of the surface point projecting to pixel \mathbf{q} at time t
- estimate the reconstruction error (Section 4.2),

$$e(r, t, \mathbf{q}) = RE(\mathbf{p}, \mathbf{n}) .$$

Step 4 Set $r^* = \arg \min_r \sum_{t, \mathbf{q}} e(r, t, \mathbf{q})$ and return the distances and normals reconstructed with this index value.

Step 5 For each time t , fuse the pixel-wise 3D position and normal estimates to obtain a 3D surface.

4. Pixel-wise Shape Estimation

The key step in refraction stereo is an optimization procedure that assigns a 3D point \mathbf{p} and a surface normal \mathbf{n} to each pixel. The procedure assumes that the refraction index has a known value r and computes the \mathbf{p}, \mathbf{n} that are most consistent with Snell’s law and the pixel-to-pattern correspondence function for the input views.

For a given pixel \mathbf{q} , the optimization works in two stages. In the first stage, we conduct a 1D optimization along the ray through pixel \mathbf{q} . The goal is to find the distance d that globally minimizes a novel criterion, called the *refractive disparity* (RD). This criterion is specifically designed to avoid instabilities due to degenerate refraction paths (e.g., when the liquid’s surface is close to the reference pattern).

The optimal d -value gives us initial estimates, $\mathbf{p}(d)$ and $\mathbf{n}(d, r)$, for the 3D coordinates and surface normal of a point that projects to pixel \mathbf{q} . These estimates are further refined in a second, bundle adjustment stage in which all five parameters (two for the normal, three for the position) are optimized simultaneously.

4.1. Measuring Refractive Disparity

Each value of d defines an implicit correspondence between four known points (Figure 1): pixel \mathbf{q} , the point $C(\mathbf{q})$ on the reference pattern that refracts to \mathbf{q} , the projection, \mathbf{q}' of $\mathbf{p}(d)$ in the second viewpoint, and point $C(\mathbf{q}')$. This correspondence must be consistent with Snell’s law.

In their work on reconstructing mirror-like surfaces, Bonfort and Sturm [22] noted that such a correspondence gives us two “candidate” normals for $\mathbf{p}(d)$ which must be identical when this hypothesis is correct. These normals are obtained by applying Eq. (2) twice, once for each viewpoint. Specifically, the first normal, $\mathbf{n}_1 = \mathbf{n}(d, r)$, ensures that point $C(\mathbf{q})$ on the reference pattern refracts to pixel \mathbf{q} via point $\mathbf{p}(d)$. The second normal, \mathbf{n}_2 , enforces a similar condition for the second viewpoint, i.e., it ensures that point $C(\mathbf{q}')$ refracts to pixel \mathbf{q}' via point $\mathbf{p}(d)$. We obtain \mathbf{n}_2 by applying Eq. (2) to pixel \mathbf{q}' , using its distance from

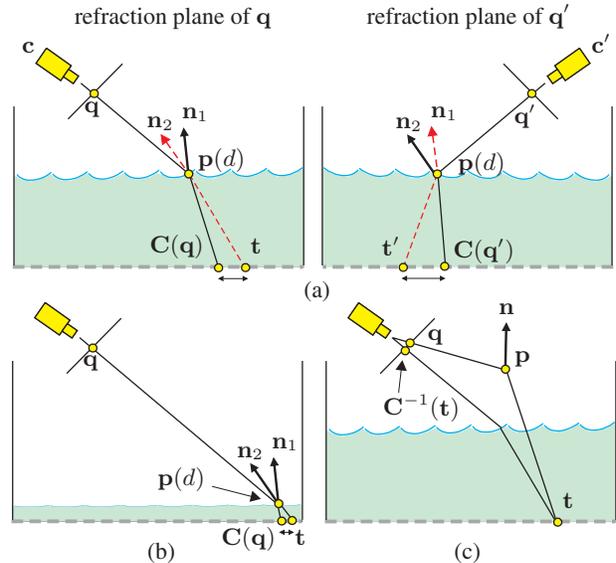


Figure 3. Optimization criteria for refraction stereo. (a) Measuring refractive disparity. Normals are drawn according to the refractions they produce. (b) For small surface-to-pattern distances, swapping \mathbf{n}_1 and \mathbf{n}_2 does not influence the distance between $C(\mathbf{q})$ and t significantly. (c) Measuring image re-projection error at one of the viewpoints.

point $\mathbf{p}(d)$. Since points on a smooth surface have a unique normal, a necessary condition for $\mathbf{p}(d)$ being on the “true” surface is that $\mathbf{n}_1 = \mathbf{n}_2$.

Unfortunately, even though it is possible, in principle, to directly measure the alignment of vectors \mathbf{n}_1 and \mathbf{n}_2 , such a measurement becomes unstable when the distance between the surface and the reference pattern approaches zero. This is because refraction diminishes, Eq. (2) becomes singular, normals cannot be estimated accurately, and the 3D reconstruction problem degenerates to standard stereo. In practice, this causes instability for low liquid heights, making direct comparison of normals uninformative and inappropriate for reconstruction.

Instead of measuring the alignment of the two normals \mathbf{n}_1 and \mathbf{n}_2 directly, we perform an indirect measurement that is not singular when refraction diminishes. The main idea is that if \mathbf{n}_1 and \mathbf{n}_2 were truly aligned, “swapping” them would still force point $C(\mathbf{q})$ to refract to pixel \mathbf{q} and point $C(\mathbf{q}')$ to pixel \mathbf{q}' . We therefore define the criterion by asking two questions (Figure 3a):

- suppose the normal at $\mathbf{p}(d)$ is \mathbf{n}_2 ; which point on the reference pattern will refract to pixel \mathbf{q} ?
- suppose the normal at $\mathbf{p}(d)$ is \mathbf{n}_1 ; which point on the reference pattern will refract to pixel \mathbf{q}' ?

Now suppose that points t, t' are the points that refract to pixels \mathbf{q}, \mathbf{q}' , respectively. The distance between t and $C(\mathbf{q})$ and, similarly, the distance between t' and $C(\mathbf{q}')$, can be thought of as a measure of disparity. Intuitively, this distance tells us how swapping the normals $\mathbf{n}_1, \mathbf{n}_2$ affects consistency with the available pixel-to-pattern correspondences. To evaluate a hypothesis d we simply sum these distances:

Refractive Disparity

$$RD(d) = \|\mathbf{t} - \mathbf{C}(\mathbf{q})\|^2 + \|\mathbf{t}' - \mathbf{C}(\mathbf{q}')\|^2. \quad (4)$$

When the distance between the true surface and the reference pattern is large, refractive disparity is equivalent to a direct measurement of the alignment between vectors $\mathbf{n}_1, \mathbf{n}_2$, i.e., it is zero if and only if $\mathbf{n}_1 = \mathbf{n}_2$. On the other hand, as the liquid’s true surface approaches the reference pattern, refractive disparity diminishes. This is because the refractive effect of changing a point’s surface orientation diminishes as well (Figure 3b). As a result, the minimization can be applied to any image pixel for which $\mathbf{C}(\mathbf{q})$ is known, regardless of whether or not the ray through the pixel actually intersects the liquid’s surface.

To compute point \mathbf{t} for a given d -value, we trace a ray from the first viewpoint through pixel \mathbf{q} , refract it at point $\mathbf{p}(d)$ according to normal \mathbf{n}_2 , and intersect it with the (known) surface of the reference pattern. Point \mathbf{t}' is computed in an identical manner. To find the distance d that globally minimizes refractive disparity along the ray we use Matlab’s `fminbnd()` function, which is based on golden section search [27].

4.2. Computing 3D Position and Orientation

Even though refractive disparity minimization yields good reconstructions in practice, it has two shortcomings. First, it treats the cameras asymmetrically, since optimization occurs along the ray through one pixel. Second, it only optimizes the distance along that ray, not the 3D coordinates and orientation of a surface point. We therefore use an additional step that adjusts all shape parameters (\mathbf{p} and \mathbf{n}) in order to minimize a symmetric image re-projection error.

To evaluate the consistency of \mathbf{p} and \mathbf{n} we check whether the refractions caused by such a point are consistent with the refractions observed in the input views. In particular, let \mathbf{q}, \mathbf{q}' be the point’s projections in the two cameras and suppose that \mathbf{t}, \mathbf{t}' are the points on the reference pattern that refract to \mathbf{q}, \mathbf{q}' , respectively, via point \mathbf{p} (Figure 3c). To compute the re-projection error we measure the distance between pixels \mathbf{q}, \mathbf{q}' and the “true” refracted image of \mathbf{t}, \mathbf{t}' :

$$RE(\mathbf{p}, \mathbf{n}) = \|\mathbf{q} - \mathbf{C}^{-1}(\mathbf{t})\|^2 + \|\mathbf{q}' - \mathbf{C}^{-1}(\mathbf{t}')\|^2 + \beta G(\|\mathbf{p} - \mathbf{p}_0\|; \sigma)^{-1}$$

where $\mathbf{C}^{-1}(\cdot)$ denotes the inverse of the pixel-to-pattern correspondence function, $G(\cdot; \sigma)$ is the Gaussian with standard deviation σ , and \mathbf{p}_0 is the starting point of the optimization. The Gaussian term ensures that \mathbf{p} is restricted to within a small neighborhood of the initial position. We used $\sigma = 4$ and $\beta = 200$ for all our experiments. To minimize the RE functional with respect to \mathbf{p} and \mathbf{n} we use the downhill simplex method [27].

5. Implementation Details

Estimating Pixel-to-Pattern Correspondences Accurate 3D shape recovery requires knowing the pixel-to-pattern correspondence function $\mathbf{C}(\mathbf{q}, t)$ with high accuracy. While color-based techniques have been used to estimate this function for image-based rendering applications [2], they are not

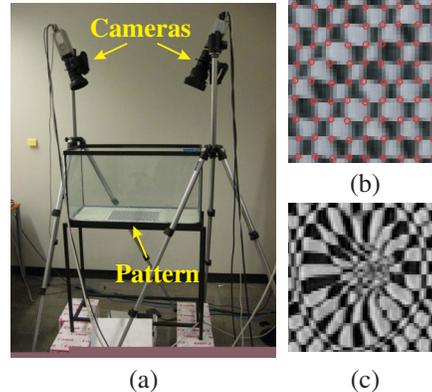


Figure 4. (a) Experimental setup. (b) Typical close-up view of pattern, seen through water surface. (c) Distorted view, corresponding to tracking failure at the central corners.

appropriate for reconstruction for several reasons. First, different liquids absorb different wavelengths by different amounts, altering a pattern’s appearance in a liquid-dependent way. Second, since light absorption depends on distance traveled within the liquid and since this distance depends on the liquid’s instantaneous shape, the appearance of the same point on a pattern will change through time. Third, the intensity of light transmitted through the surface depends on the Fresnel effect [7] and varies with wavelength and the angle of incidence. This makes it difficult to use color as a means to localize points on a pattern with near- or sub-pixel accuracy.

In order to avoid these complications, we use a monochrome checkered pattern and rely on corners to establish and maintain pixel-to-pattern correspondences (Figures 4a,b). We assume that the liquid’s surface is undisturbed at time $t = 0$ and use the Harris corner detector [28] to detect corners at sub-pixel resolution. This gives us the initial pixel-to-pattern correspondences. To track the location of individual corners in subsequent frames while avoiding drift, we estimate flow between the current frame and the frame at time $t = 0$, using the flow estimates from the previous frame as an initial guess. We compute flow with a translation-only version of the Lucas-Kanade inverse-compositional algorithm, with Levenberg-Marquardt minimization to give sub-pixel registration [29]. This algorithm is applied to an 11×11 pixel neighborhood around each corner. We use the registration error returned by the algorithm as a means to detect failed localization attempts. In the case of failure, the flow computed for that corner is not used and the corner’s previously known location is propagated. This allows our tracker to overcome temporary obscurations due to blur, splashes or extreme refractive distortions (Figure 4c). The above procedure gives values of the correspondence function $\mathbf{C}(\mathbf{q}, t)$ for a subset of the pixels. To evaluate the function for every pixel, we use bilinear interpolation.

Fusing 3D Positions and Orientations Refraction stereo yields a separate 3D position and 3D normal for each pixel. While this is a richer shape descriptor, the problem of reconstructing a single surface that is consistent with both

types of data is still open. A key difficulty is that point and normal measurements have different noise properties and, hence, a surface computed via normal integration and a surface computed by fitting a mesh to the 3D points will not necessarily agree. As a first step, we used simulations and ground-truth experiments to estimate the reliability of each data source as a function of surface height, i.e., distance from the plane of the reference pattern (Figure 5). Since reconstructed normals are highly reliable for large heights, we used this analysis to set a height threshold, below which normals are deemed less reliable than positions. That portion of the surface is reconstructed from positional data. For the remaining pixels, we reconstruct the surface via normal integration, using the Ikeuchi-Horn algorithm [30], and merge the results. In cases where all reconstructed positions are above the height threshold, we rely on normal integration to compute 3D shape and use the average 3D position to eliminate the integrated surface’s height ambiguity.

6. Experimental Results

Experimental setup Figure 4a shows our setup. The checkered pattern at the bottom of the tank was in direct contact with the water to avoid secondary refractions. During our experiments, the pattern was brightly lit from below to avoid specular reflections and to enable use of a small aperture size for the cameras (and, hence, a large depth of field). Images were acquired at a rate of 60Hz with a pair of synchronized Sony DXC-9000 progressive-scan cameras, whose electronic shutter was set to $1/500sec$ to avoid motion blur. Both cameras were approximately 1 meter above the tank bottom and were calibrated using the Matlab Calibration Toolbox [31].

Simulations To evaluate the stability of our algorithms, we performed simulations that closely matched the experimental conditions in the lab (e.g., relative position of cameras, pattern-to-camera distances, feature localization errors, etc.). We simulated the reconstruction error for planar water surfaces as a function of the surface-to-pattern height, and for various levels of error in corner localization and camera calibration. For each height, we reconstructed 10,000 individual points and measured their deviation from the ground-truth plane (Figure 5). These simulations confirm that the accuracy of reconstructed normals degrades quickly for water heights less than $4mm$. Importantly, the accuracy of distance computations is not sensitive to variations in water height, confirming the stability of our optimization-based framework for refractive stereo.

Accuracy experiments Since ground truth was not available, we assessed our algorithm’s accuracy by applying it to the reconstruction of flat water surfaces whose height from the tank bottom ranged from 4 to $15mm$. For each water height, we reconstructed a point \mathbf{p} and a normal \mathbf{n} independently for each of 1,836 pixels in the two image planes, giving rise to as many 3D points and normals. No smoothing or post-processing was performed. To assess the accuracy of the reconstructed points, we fit a plane using least squares and measured the points’ RMS distance from

this plane. To assess accuracy in the reconstructed normals, we computed the average normal and measured the mean distance of each reconstructed normal from the average normal. These results, also shown in Figure 5, closely match the behavior predicted by our simulations. They also suggest that reconstructions are highly precise, with distance variations around $0.25mm$, (i.e., within 99.97% of the surface-to-camera distance) and normal variations on the order of 2 degrees for water heights above $8mm$.

Experiments with dynamic surfaces Figure 6 and supplementary videos in [32, 33] show reconstructions of several dynamic water surfaces. The experiments test our algorithm’s capabilities under a variety of conditions, from rapidly-fluctuating water that is high above the tank bottom, to water that is being poured in an empty tank, where the water height is very small and refraction is degenerate or near-degenerate for many pixels. Figure 5(right) shows the total reconstruction error corresponding to specific values of the refractive index, for the “waves” dataset [32, 33]. While the curve does not exhibit a very steep minimum, the minimum is obtained at the correct value for water, 1.33, confirming the predictions of Theorem 1. Several observations can be made from these experiments. First, our tracking-based framework allows us to maintain accurate pixel-to-pattern correspondences for 100s of frames, enabling dynamic reconstructions that last several seconds. Second, the reconstructed distances remain stable despite large variations in water height, and are accurate enough to show fine surface effects even in cases where the total water height never exceeds $6mm$ (e.g., the “pour” sequence). Third, the reconstructed normal maps, as predicted, show fine surface fluctuations at larger heights but degrade to noise levels for water heights near zero. Qualitatively, when there is sufficient water in the tank, they appear to contain less noise than depth maps. Fourth, the normal integration algorithm that we are currently using seems to over-smooth fine surface details that are clearly present in the depth and normal maps. Hence, the question of how to best extract surfaces from the raw data provided by refraction stereo is still open. Fifth, our search-based framework for refractive index estimation produces stable, unambiguous results because it enforces a very strong global constraint: a single refractive index value must account for the refractions caused by the *entire sequence* of 3D surfaces of the deforming liquid, not just the 3D surface at a single instant.

7. Concluding Remarks

Liquids can generate extremely complex surface phenomena, including breaking waves, bubbles, and extreme surface distortions. While our refraction stereo results are promising, they are just an initial attempt to model liquid flow in relatively simple cases. Our ongoing work includes (1) reconstructing surfaces that produce multiple refractions [34], (2) reconstructing liquids by exploiting their refractive and reflective properties (e.g., by also treating them as mirrors), and (3) “reusing” captured 3D data to create new, realistic fluid simulations.

Acknowledgements This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under the RGPIN program, by the Province of Ontario under the OGSST program, by a fellowship from the Alfred P. Sloan Foundation, and by an Ontario Premier’s Research Excellence Award.

Proof sketch of Theorem 1 The proof uses two basic intuitions. First, given an arbitrary value for the refractive index, each viewpoint can be thought of as defining a *3D constraint curve*, representing all assignments of distances and normals to a pixel \mathbf{q} that are compatible with Snell’s law. Hence, an assignment that is consistent with both viewpoints corresponds to the intersection of two such constraint curves. Second, and most important, for an arbitrarily-shaped surface, these 3D curves will be in general position with respect to each other and, therefore, will not have a common intersection. From these two facts we conclude that when the refractive index has an arbitrary value, there will be no distance and normal assignment that is consistent with both viewpoints. Hence, such consistency can only be achieved for isolated refractive index values. We formalize these intuitions below.

Let r^* be the true refractive index of the surface and let $r \neq r^*$ be an arbitrary value of this index. Without loss of generality, we assume that the function $\mathbf{C}(\mathbf{q})$ is known for all pixels \mathbf{q} .

Let \mathbf{q} be an arbitrary pixel in the first viewpoint and let d^* be its true distance to the surface. Given value r for the refractive index, every distance d defines a unique normal, $\mathbf{n}(d, r)$, compatible with Snell’s law (Figure 2 and Eq. (2)). Now suppose we represent unit vectors with two angles: an angle θ , corresponding to the angle between the vector and the ray through \mathbf{q} ; and an angle ϕ , corresponding to the angle between the vector and the normal of \mathbf{q} ’s refraction plane. In this representation, the distance and normal assignments to \mathbf{q} that are compatible with Snell’s law define a curve γ in (d, θ, ϕ) -space. This curve will always lie on the plane $\phi = 0$ since, by definition, the normal $\mathbf{n}(d, r)$ always lies on the refraction plane of pixel \mathbf{q} .

Now let $\mathbf{q}'(d)$ be the projection of $\mathbf{p}(d)$ in the second viewpoint (Figure 1). Since $\mathbf{C}(\mathbf{q}'(d))$ is known, there is only one normal, $\mathbf{n}'(d, r)$, that can be assigned to $\mathbf{p}(d)$ and is compatible with Snell’s law in the second viewpoint. This normal will lie on the refraction plane of pixel \mathbf{q}' . Generically, the refraction planes of pixels \mathbf{q} and \mathbf{q}' are distinct. Hence, the normal $\mathbf{n}'(d, r)$ may not lie on the refraction plane of pixel \mathbf{q} and, as d varies, $\mathbf{n}'(d, r)$ will trace a general curve γ' in (d, θ, ϕ) -space, i.e., a curve that is not restricted to the plane $\phi = 0$.

We now show that γ and γ' do not intersect. First note that the two curves cannot intersect in the neighborhood of the “true” distance d^* because $\mathbf{n}'(d^*, r) \neq \mathbf{n}(d^*, r)$.² Now consider distances away from d^* . We show that γ' and γ generically will not intersect there either. In particular, the normal $\mathbf{n}'(d, r)$ is completely determined by point $\mathbf{C}(\mathbf{q}'(d))$ which, in turn, is determined by the normal of the true surface point projecting to pixel $\mathbf{q}'(d)$. Since $\mathbf{q}'(d)$ lies on the epipolar line of \mathbf{q} for all values of d , it follows that curve γ' is completely determined by the surface normal of points at the intersection, C , of this epipolar plane with the true surface. For γ' and γ to intersect there must be a point on C outside the neighborhood of $\mathbf{p}(d^*)$ whose surface normal is identical to $\mathbf{p}(d^*)$. This condition, however, cannot be satisfied for an open 2D set of points on a generic surface. It follows that γ' and γ are

²Observe that $\mathbf{n}'(d^*, r^*) = \mathbf{n}(d^*, r^*)$ since Snell’s law is satisfied for both viewpoints in the true scene. Now, since there is a 1-1 correspondence between refractive indices and normals when d^* is fixed, and since the refraction planes of \mathbf{q} and $\mathbf{q}'(d)$ have only one normal in common, it follows that $\mathbf{n}'(d^*, r) \neq \mathbf{n}(d^*, r)$.

non-intersecting for almost all points on the surface and, hence, for almost all pixels in the surface’s projection.

References

- [1] D. Enright, S. Marschner, and R. Fedkiw, “Animation and rendering of complex water surfaces,” *Proc. SIGGRAPH’02*, pp. 736–744.
- [2] Y.-Y. Chuang, D. E. Zongker, J. Hirdorff, B. Curless, and D. H. Salesin, “Environment matting extensions: Toward higher accuracy and real-time capture,” *Proc. SIGGRAPH’00*, pp. 121–130.
- [3] F. Harlow and J. Welch, “Numerical calculation of time-dependent viscous incompressible flow,” *Phys. Fluids*, v. 8, pp. 2182–2189, 1965.
- [4] B. Jähne, J. Klinke, S. Waas, “Imaging of short ocean wind waves: a critical theoretical review,” *JOSA-A*, v. 11, pp. 2197–2209, 1994.
- [5] W. C. Keller and B. L. Gotwols, “Two-dimensional optical measurement of wave slope,” *Applied Optics*, v. 22, pp. 3476–3478, 1983.
- [6] L. Zhang, N. Snavely, B. Curless, and S. M. Seitz, “Spacetime faces: high resolution capture for modeling and animation,” *Proc. SIGGRAPH’04*, pp. 548–558.
- [7] A. S. Glassner, *Principles of Digital Image Synthesis*. Morgan Kaufmann, 1995.
- [8] L. Zhou, C. Kambhamettu, and D. B. Goldgof, “Fluid structure and motion analysis from multi-spectrum 2D cloud image sequences,” *Proc. CVPR’00*, v. 2, pp. 744–751.
- [9] J. Höhle, “Reconstruction of the underwater object,” *Photogrammetric Engineering*, pp. 948–954, 1971.
- [10] H.-G. Maas, “New developments in multimedia photogrammetry,” *Optical 3D Measurement Techniques III*, Wichmann Verlag, 1995.
- [11] P. Flach and H.-G. Maas, “Vision-based techniques for refraction analysis in applications of terrestrial geodesy,” *Int. Archives of Photogrammetry and Remote Sensing*, pp. 195–201, 2000.
- [12] A. Okamoto, “Orientation problem of two-media photographs with curved boundary surfaces,” *Photogrammetric Engineering and Remote Sensing*, pp. 303–316, 1984.
- [13] H. Murase, “Surface shape reconstruction of an undulating transparent object,” *Proc. ICCV’90*, pp. 313–317.
- [14] S. Agarwal, S. P. Mallick, D. Kriegman, and S. Belongie, “On refractive optical flow,” *Proc. ECCV’04*, pp. 483–494.
- [15] M. Tarini, H. P. A. Lensch, M. Goesele, and H.-P. Seidel, “3D acquisition of mirroring objects,” Tech. Rep. MPI-I-2003-4-001, Max-Planck-Institut für Informatik, 2003.
- [16] X. Zhang and C. S. Cox, “Measuring the two-dimensional structure of a wavy water surface optically: A surface gradient detector,” *Experiments in Fluids*, vol. 17, pp. 225–237, 1994.
- [17] A. Blake, “Specular stereo,” *Proc. IJCAI*, pp. 973–976, 1985.
- [18] M. Oren and S. K. Nayar, “A theory of specular surface geometry,” *Proc. ICCV’95*, pp. 740–747.
- [19] S. Savarese and P. Perona, “Local analysis for 3d reconstruction of specular surfaces - part ii,” *Proc. ECCV’02*, pp. 759–774.
- [20] K. Ikeuchi, “Determining the surface orientations of specular surfaces by using the photometric stereo method,” *IEEE T-PAMI*, v. 3, n. 6, pp. 661–669, 1981.
- [21] A. Sanderson, L. Weiss, and S. Nayar, “Structured highlight inspection of specular surfaces,” *IEEE T-PAMI*, v. 10, n. 1, pp. 44–55, 1988.
- [22] T. Bonfort and P. Sturm, “Voxel carving for specular surfaces,” *Proc. ICCV’03*, pp. 591–596.
- [23] R. I. Hartley and P. Sturm, “Triangulation,” *Computer Vision and Image Understanding: CVIU*, v. 68, n. 2, pp. 146–157, 1997.
- [24] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, “Bundle adjustment - a modern synthesis,” *Proc. Int. Workshop on Vision Algorithms*, pp. 298–372, 2000.
- [25] T. Zickler, P. N. Belhumeur, D. J. Kriegman, “Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction,” *Proc. ECCV’02*, pp. 869–884.
- [26] J. J. Koendering and A. J. van Doorn, “The structure of two-dimensional scalar fields with applications to vision,” *Biological Cybernetics*, v. 33, pp. 151–158, 1979.
- [27] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C*. Cambridge University Press, 1988.
- [28] C. Harris and M. Stephens, “A combined edge and corner detector,” *Proc. 4th Alvey Vision Conference*, pp. 189–192, 1988.
- [29] S. Baker and I. Matthews, “Lucas-kanade 20 years on: A unifying framework,” *IJCV*, v. 56, n. 3, pp. 221–255, 2004.
- [30] K. Ikeuchi and B. K. P. Horn, “Numerical shape from shading and occluding boundaries,” *Artificial Intelligence*, v. 17, pp. 141–184, 1981.

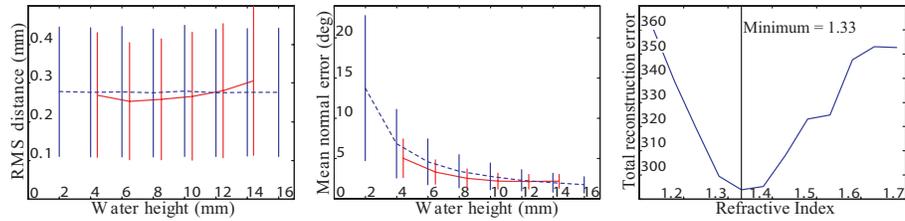


Figure 5. **Left & middle:** Reconstruction accuracy as a function of water height, for real (solid line) and simulated (dotted line) flat water surfaces. Bars indicate standard deviation. Simulations are for a 0.08-pixel localization error; in real flat water experiments, corner localization precision was measured to be 0.1 pixels. **Right:** Total reconstruction error as a function of refractive index.

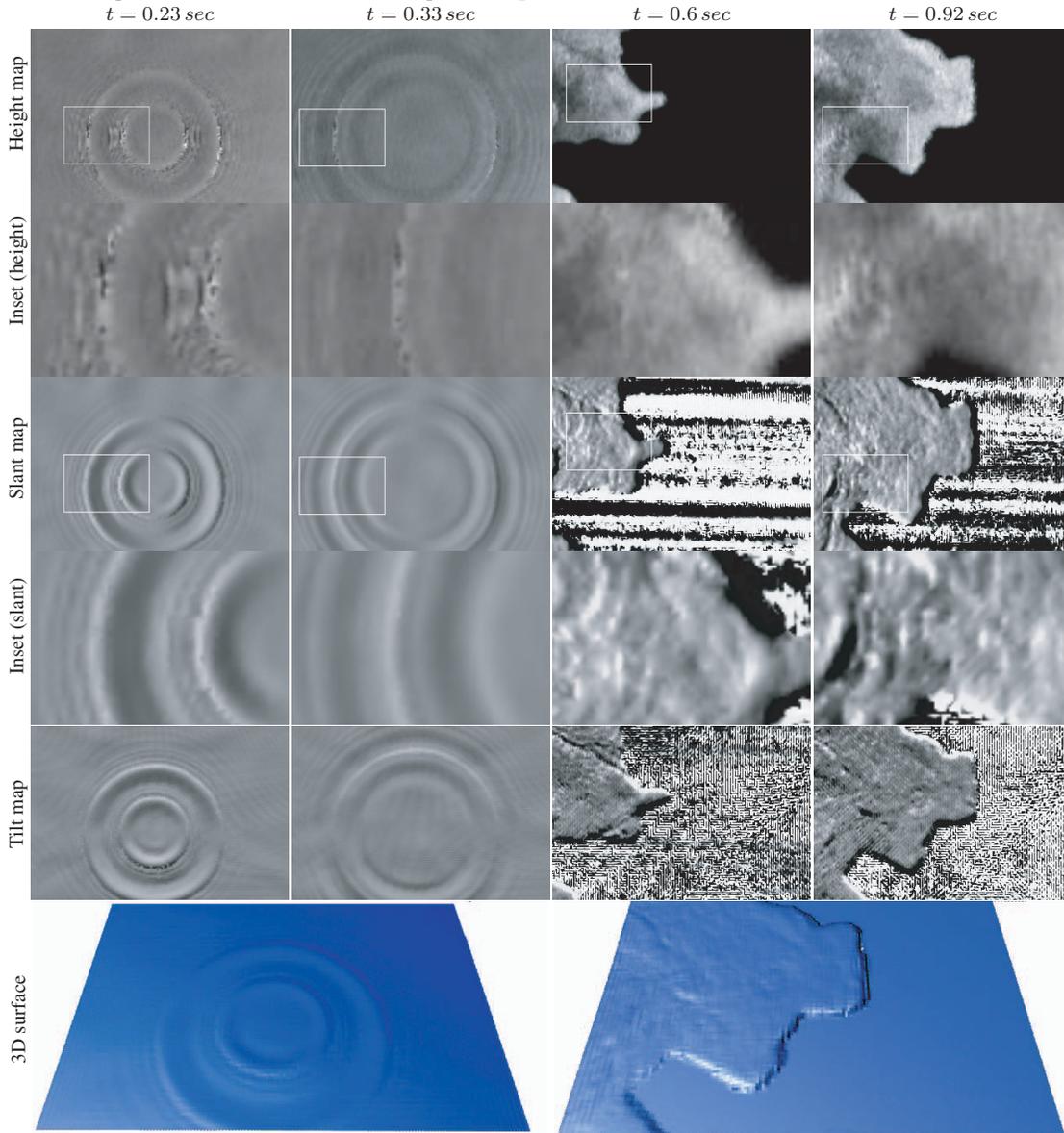


Figure 6. 3D reconstruction results. All maps correspond to a top view of the tank and show raw, per-pixel data. Height grayscale values are mapped to $[21\text{mm}, 33\text{mm}]$ for the "ripple" sequence (left) and to $[0\text{mm}, 6\text{mm}]$ for the "pour" sequence. Normal maps represent the slant and tilt angle of each reconstructed normal and range from -90° (black) to 0 (gray) to $+90^\circ$ (white). See the supplementary videos in [32, 33] for more information on these sequences and for more examples.

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

[32] N. J. W. Morris, "<http://www.dgp.toronto.edu/~nmorris/Water>".

[33] ICCV'05 Proceedings CD-ROM.

[34] K. N. Kutulakos and E. Steger, "A Theory of Specular and Refractive Shape by Light-Path Triangulation," *Proc. ICCV'05*.