Stereo-Based 3D Reconstruction of Dynamic Fluid Surfaces by Global Optimization

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

3D reconstruction of dynamic fluid surfaces is an open and challenging problem in computer vision. Unlike previous approaches that reconstruct each surface point independently and often return noisy depth maps, we propose a novel global optimization-based approach that recovers both depths and normals of all 3D points simultaneously. Using the traditional refraction stereo setup, we capture the wavy appearance of a pre-generated random pattern, and then estimate the correspondences between the captured images and the known background by tracking the pattern. Assuming that the light is refracted only once through the fluid interface, we minimize an objective function that incorporates both the cross-view normal consistency constraint and the single-view normal consistency constraints. The key idea is that the normals required for light refraction based on Snell's law from one view should agree with not only the ones from the second view, but also the ones estimated from local 3D geometry. Moreover, an effective reconstruction error metric is designed for estimating the refractive index of the fluid. We report experimental results on both synthetic and real data demonstrating that the proposed approach is accurate and shows superiority over the conventional stereo-based method.

1. Introduction

The problem of 3D reconstruction of dynamic fluid surfaces has attracted much attention from many areas, including computer vision [19], oceanology [14] and computer graphics [8]. Effective solutions can benefit many applications, *e.g*. physics-based fluid simulation [10] and fluid imaging [2]. The problem is challenging for several reasons. First, similar to glass and crystal, most fluids are transparent and do not have their own colors. They acquire their colors from surrounding backgrounds. Hence, traditional color-based stereo matching cannot work for such view-dependent surfaces. Second, tracing the light path involved in fluid surface reconstruction is non-trivial because of the non-linearity inherent in refraction. Even worse is that light refraction depends not only on the 3D shape but also on the medium's property, *i.e*. refractive index, which is usually unknown. Third, compared to static transparent objects, accurately reconstructing wavy fluid surfaces is even harder because real time capture is required.

In computer vision, the problem is usually solved via shape from refraction. Typically, a known background is placed beneath the fluid surface and 3D reconstruction is performed by analyzing pixel-point correspondences. That is, for each pixel, the corresponding location of the light source in the background is acquired. However, shape from pixel-point correspondence is known to have ambiguities: the 3D surface point can lie at any position along the camera ray that goes through the pixel. Recent methods resolve the ambiguities along two directions. Some methods [29, 31, 34], instead of using pixel-point correspondences, acquire *ray-ray* correspondences, *i.e*. the incident ray emitted from the background and the exit ray going to the camera, using special devices (*e.g*. Bokode [31], light field probes [29]). Alternatively, a number of methods [7, 19] propose to employ stereo/multiple cameras to capture the fluid surface, which basically utilize a cross-view normal consistency constraint: the normals computed using the pixel-point correspondences acquired from different views should be consistent. Nevertheless, for the above two groups, a common limitation is that they result in reliable normals only but noisy depths. The final 3D points of the fluid surface are then obtained by normal integration. To get the boundary condition for integration, they either assume that the surface is flat at the boundary [7, 31] or estimate the boundary using the noisy depths [19, 29].

To cope with the above limitations, we propose a global optimization-based approach to reconstruct a dynamic, homogeneous and transparent fluid surface, from which specular reflection is assumed to be negligible. Our approach is based on pixel-point correspondences. By assuming light is redirected only once through the fluid surface, we first use two perspective cameras to capture the distortion of a ran-

dom pattern through the wavy surface. Hence, our acquisition system is easy to implement and requires no special optics. Compared to a conventional stereo-based method [19], the proposed approach can obtain both accurate, consistent depths and normals without the error-prone surface integration step. Specifically, rather than doing a point-by-point reconstruction, we formulate a global optimization function, which exploits not only the cross-view normal consistency but also the single-view normal consistency constraints. By doing so, we jointly reconstruct depths and normals. Our method addresses the fundamental limitation of existing methods on surface integration without accurate boundary conditions. Besides, a new reconstruction error metric is designed to search the refractive index of liquid with very encouraging results.

2. Related Work

The **single-view** based method was first introduced by Murase [20] in computer vision, where surface normals are recovered by capturing video with an orthographic camera of a flat background through wavy water. To eliminate the ambiguity in pixel-point correspondences, earlier efforts focus on proposing additional constraints, *e.g*. statistical appearance assumption of a fluid sequence [20], known average fluid height [14]. Recently, Shan *et al*. [25] improve Murase's method by solving all surface points at the same time under orthographic projection. However, their implementation requires a long exposure time (about 0.5 seconds) for each frame and thus is applicable to static objects only. By modeling the surface as a cubic B-spline, Liu *et al*. [18] introduce a parametric solution for reconstructing both mirror objects and transparent surfaces using pixel-point correspondences.

Ray-ray correspondence based methods are developed to avoid the ambiguity of pixel-point correspondences under a single-view setup. By placing a color screen at the focal length of a big lens, Zhang and Cox [34] associate each 2D source point of the background with a ray direction under orthographic projections. The incident rays are then easily obtained after getting pixel-point correspondences. Ye *et al*. [31] establish a similar setup by using a perspective camera. Wetzstein *et al*. [29] acquire ray-ray correspondences with light field probes [28]. Specifically, they replace the big lens with a lenslet array. A color pattern is then placed under the array, which encodes positional and angular correspondences using different color channels. All the above ray-ray correspondence based methods rely on special optics, which introduces many practical issues, *e.g*. calibrating the ray directions of background points [15] and making the setup waterproof [31]. In addition, as reported in their papers [29, 31], the surface positions obtained by intersecting the incident and exit rays are less accurate than that of the normals obtained by Snell's law. Furthermore, a surface integration algorithm is required to obtain the 3D shape from the normal information.

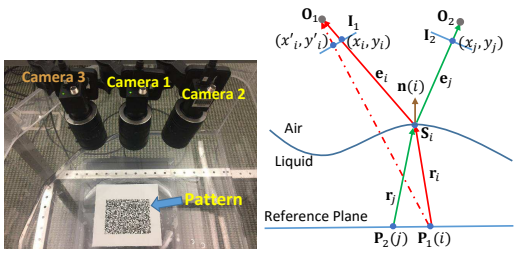
Another group of methods utilize **multiple viewpoints** to tackle the problem. Morris and Kutulakos [19] first propose using a stereo camera system to capture a dynamic fluid surface. By placing a checkerboard underneath the fluid surface, their approach can estimate both depths and normals based on pixel-point correspondences. Following their stereo setup, our approach not only inherits the advantage of easy implementation (*e.g*. no special devices required and can work under perspective projection) but also provides the following novel improvements: (1) In addition to cross-view normal consistency, our approach exploits a novel single-view normal consistency which takes local surface geometry into account; (2) Unlike their method which solves for each individual point independently, ours employs a global optimization scheme to recover all surface points simultaneously which results in higher accuracy in both depth and normal; (3) Since they compute depths and normals in separate steps, the surface obtained by mesh fitting based on the depth map and the one estimated via normal integration do not guarantee consistency. Typically, their normals are more accurate than the corresponding depths. Thus an additional surface integration from normals is required. In comparison, we simultaneously reconstruct depths and normals, which are both accurate and, most importantly, are consistent with each other; (4) We define a new error metric to recover the unknown refractive index without requiring to compute the complex inverses of pixel-point correspondences as in their method. It is noteworthy that the refraction stereo formulation has been extended to using a camera array [7], where the fluid surface is reconstructed by specular carving. However, the major limitations of [19] discussed above remain unsolved.

3D fluid surfaces can also be recovered based on light reflections [17, 32]. In addition, our work is also closely related to the problem of reconstructing static transparent objects [11, 16, 22, 26, 37] and dynamic gas flows [3, 15, 30]. Interested readers are referred to the surveys [12, 13] of this field.

3. Proposed Approach

3.1. Correspondence Acquisition and Matching

Our approach computes the 3D shape of a transparent fluid surface based on how it refracts light. Specifically, for each pixel, the position of the corresponding background point is required, *i.e*. pixel-point correspondence. As shown in Fig. 1(a), we place a pre-generated pattern at the bottom of a tank, and capture the scene from two different viewpoints with Camera 1 and Camera 2, respectively. For each camera, we first capture the pattern without water as a reference image $\mathbf{B}$. The cameras are synchronized for capturing



(a) (b)

Figure 1. Acquisition setup (a) and the corresponding refraction stereo geometry (b). Note that Camera 3 in the left figure is for accuracy assessment only and not used during 3D reconstruction.

dynamic surfaces after adding water. Note that an additional camera (Camera 3) is used for accuracy evaluation using an image-based rendering method in our real experiments, which is discussed in Sec. 4.2.

Fig. 1(b) illustrates the acquisition setup in 2D. Consider two perspective cameras centered at $\mathbf{O_1}$ and $\mathbf{O_2}$ observing a refractive surface against a flat background. Taking Camera 1 for example, for each pixel (x_i, y_i) in Camera 1, light originating from the corresponding point $\mathbf{P}_1(i)$ on the reference plane gets refracted at surface point $\mathbf{S}_i$. Let $\mathbf{I}_1(x_i, y_i, t)$ be the tth captured frame of the refraction distorted pattern. Here, the goal is to estimate the light source point $\mathbf{P}_1(i)$ for each pixel (x_i, y_i). We first apply the coarse-to-fine variational framework [5] to compute the optical flow (u_i, v_i) between $\mathbf{I}_1$ and $\mathbf{B_1}$. Then the forward projection (x'_i, y'_i) of point $\mathbf{P}_1(i)$ is easily computed as $(x_i + u_i, y_i + v_i)$. Suppose the relative poses between the cameras and the reference plane are calibrated beforehand and fixed during acquisition, the 3D coordinates of $\mathbf{P}_1(i)$ is estimated by intersecting ray $\mathbf{O}_1\mathbf{P}_1(i)$ with the reference plane.

The choice of the displayed pattern is critical for accurate correspondence matching and subsequent 3D reconstruction. Traditional methods [7, 19] use a checkerboard pattern and track the feature corners. Correspondences of non-corner pixels are obtained by interpolation. However, these methods assume that the first frame of the liquid surface is nearly flat, which is usually impractical, so that a reliable initial correspondence field can be obtained. In contrast, inspired from the successful applications of random patterns in single-shot structured light [9], we choose a binary random pattern generated from Bernoulli distributions [27] as shown in Fig. 9. Different from a regular checkerboard, a Bernoulli random pattern contains fewer repetitive structures, which helps to reduce ambiguities while searching correspondences in a local window. Besides, the binary random pattern extends the advantage of a checkerboard in handling light absorption, dispersion, chromatic abberations, etc, compared to color-based ones [6, 29].

The correspondence matching for Camera 2 works analogously. The same procedure is applied to different frames. So far, we have obtained the pixel-point correspondences of a liquid motion sequence from two cameras. Next, we present a novel reconstruction framework that solves the following problem: *Given the pixel-point correspondence function $\mathbf{P}_1()$ and $\mathbf{P}_2()$ of each frame from two views, how to recover the depths and the normals of the dynamic surface, as well as the refractive index?*

3.2. Stereo-Based Reconstruction

Our approach formulates a global optimization framework which enforces two forms of normal consistency constraints. Specifically, for each 3D point, the normals estimated based on light refraction from two different viewpoints should be consistent. On the other hand, they are also required to agree with the normal estimated based on single-view local shape geometry.

3.2.1 Normal Definitions

Here we first explain the definitions of the different types of normals mentioned above. Similar to color-based stereo matching, we set Camera 1 as the primary camera and the fluid surface is represented by a depth[1] map $\mathbf{D}$ in the scope of Camera 1. As shown in Fig. 1(b), for the ith surface point $\mathbf{S}_i$ associated with pixel (x_i, y_i) of Camera 1, let d_i be its hypothesized depth. The 3D coordinates of $\mathbf{S}_i$ can then be computed by first assuming that the camera's parameters are known. Given the pixel-point correspondence $\mathbf{P}_1(i)$, we get the ray direction $\mathbf{r}_i$ by connecting $\mathbf{P}_1(i)$ and $\mathbf{S}_i$. Then, the normal of $\mathbf{S}_i$ can be computed based on Snell's law, given the incident and exiting rays $\mathbf{r}_i$ and $\mathbf{e}_i$, respectively. We refer to this normal as the *LeftSnell* normal, denoted by $\mathbf{n}_1(i)$. Snell's law states that the normal $\mathbf{n}_1(i)$, the incident ray $\mathbf{r}_i$ and the exiting ray $\mathbf{e}_i$ are co-planar, and thus $\mathbf{n}_1(i)$ can be represented as a linear combination of $\mathbf{r}_i$ and $\mathbf{e}_i$. That is, $\mathbf{n}_1(i) = (\eta_l \mathbf{r}_i - \eta_a \mathbf{e}_i)/\|\eta_l \mathbf{r}_i - \eta_a \mathbf{e}_i\|$, where η_l and η_a denote the refractive index of liquid and air, respectively. We set $\eta_a = 1$ in our experiments and here the medium's refractive index η_l is assumed to be known. How to deal with fluid surface with an unknown refractive index is discussed in Sec. 3.3.

On the other hand, by connecting $\mathbf{S}_i$ and $\mathbf{O}_2$, we get ray $\mathbf{e}_j$ and the forward projection (x_j, y_j). Similarly, since the correspondence source function $\mathbf{P}_2(j)$ is acquired beforehand, we can also compute another normal of $\mathbf{S}_i$ by Snell's law given light rays $\mathbf{r}_j$ and $\mathbf{e}_j$. We refer to this normal as the *RightSnell* normal, denoted by $\mathbf{n}_2(i)$. In a similar vein, $\mathbf{n}_2(i)$ is estimated by $\mathbf{n}_2(i) = (\eta_l \mathbf{r}_j - \eta_a \mathbf{e}_j)/\|\eta_l \mathbf{r}_j - \eta_a \mathbf{e}_j\|$.

[1] In this paper, depth is defined as the distance between a 3D point and the camera center along the z axis.

In addition, the normal of a 3D point can be computed from its local shape geometry. That is, from the 3D locations of the neighboring points of $\mathbf{S}_i$, we can fit a tangent plane. Then the normal of $\mathbf{S}_i$ is approximated by the normal of the tangent plane. In particular, we estimate this normal by Principal Component Analysis (PCA) [24], which is referred to as the *PCA* normal and denoted by $\mathbf{n}_p(i)$. The basic idea is to analyze the eigenvectors and eigenvalues of a covariance matrix constructed from nearby points of the query point. More specifically, the covariance matrix $\mathcal{M}$ at the point $\mathbf{S}_i$ is defined as:

$$\mathcal{M} = \frac{1}{|\mathcal{N}(i)|} \sum_{k \in \mathcal{N}(i)} (\mathbf{S}_k - \mathbf{S}_i)(\mathbf{S}_k - \mathbf{S}_i)^\top, \quad (1)$$

where $\mathcal{N}(i)$ denotes the local neighborhood of pixel i and $|\mathcal{N}(i)|$ the size of $\mathcal{N}(i)$. The *PCA* normal $\mathbf{n}_p(i)$ is thus the eigenvector of $\mathcal{M}$ with minimal eigenvalue.

3.2.2 Objective Function

To this end, we obtain three different normal estimations computed from different sources for each surface point $\mathbf{S}_i$. Ideally, the three estimates should be the same. Therefore, the difference between each pair of normals can be used to defined a normal consistency error. That is:

$$E_{12}(i) = 1 - \mathbf{n}_1(i) \cdot \mathbf{n}_2(i), \quad (2)$$
$$E_{1p}(i) = 1 - \mathbf{n}_1(i) \cdot \mathbf{n}_p(i), \quad (3)$$
$$E_{2p}(i) = 1 - \mathbf{n}_2(i) \cdot \mathbf{n}_p(i), \quad (4)$$

where E_{12} measures the cross-view normal consistency error, which is the one used in [19]. E_{1p} and E_{2p} are our new single-view normal consistency errors.

Furthermore, assuming that the fluid surface is piecewise smooth, we define the depth smoothness term at the ith point as:

$$E_{so}(i) = \sum_{k \in \mathcal{G}(i)} (d_i - d_k)^2, \quad (5)$$

where $\mathcal{G}(i)$ is the neighborhood pixel set containing the bottom and the right pixel of pixel i in our implementation.

Summing the above error terms and considering all the surface points, we obtain the following global minimization problem:

$$\min_{d_i \in \mathbf{D}} \sum_{i \in \Omega_1} \left(\alpha E_{1p}(i) + \beta E_{2p}(i) + \gamma E_{12}(i) + \lambda E_{so}(i)\right), \quad (6)$$

where Ω_1 denotes the pixel set containing all the surface points in the region of interest. Hence, Eq.(6) couples both cross-view and single-view normal consistency constraints to optimize for the depths of all points simultaneously, whereas previous methods [7, 19] consider the cross-view error term E_{12} only and solve for each point independently. α, β, γ and λ are the parameters balancing different terms.

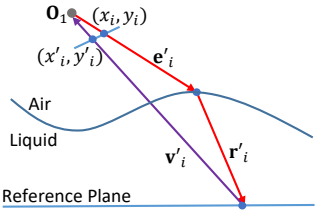


Figure 2. Ray tracing geometrically to estimate the shape-based optical flow field.

Note that Eq.(6) is defined w.r.t. a single frame. It is possible to solve the depth maps of all points from all frames by including them in Eq.(6) at the same time, which yield a large system that is computationally expensive. In contrast, we solve each frame independently and use the result of the last frame to initialize the current frame, which not only drastically reduces the running time and memory consumption but also maintains temporal coherence.

In addition, because of the complex operations involved computing the three normals, it is difficult to analytically derive the derivatives of Eq.(6). To tackle that, the previous method [19] employs the gold-section search [21] for pixelwise 1D optimization, which is computationally intensive when the number of unknowns is large and thus, the method is not applicable to our global objective function. Instead, in our implementation, we use the L-BFGS-B [36] method to optimize Eq.(6) using numerical differentiation.

3.3. Optimizing Depths and Refractive Index

As mentioned in Sec. 3.2.1, computing the *LeftSnell* and *RightSenll* normals both require the refractive index of the fluid. Given different refractive index hypotheses, solving Eq.(6) returns different depth maps. Hence, additional steps are required to get the desired 3D model when the index is unknown. Following previous methods [19, 22, 25], here we use a brute-force search approach. That is, we enumerate possible index hypotheses, evaluate the corresponding models based on a novel reconstruction error metric and pick the index with the minimal residual error.

The main idea of our proposed reconstruction error metric is based on the consistency of two optical flow fields estimated using different methods. On the one hand, as introduced in Sec. 3.1, for the ith pixel in Camera 1, we can compute the displacement vector (u_i, v_i) between the fluid image $\mathbf{I}_1$ and the reference image $\mathbf{B}_1$ using image-based cues [5]. On the other hand, since the 3D shape of the fluid surface is reconstructed, the flow field can also be obtained using shape-based cues. As shown in Fig. 2, for the ith pixel (x_i, y_i), we trace along each camera ray $\mathbf{e}'_i$ and locate

its intersection with the fluid surface. The refracted ray $\mathbf{r}'_i$ is then obtained by Snell's law. Finally, the pixel coordinates (x'_i, y'_i) are obtained by projecting back to the camera center along the direction $\mathbf{v}'_i$, and the shape-based displacement vector is computed as $(u'_i, v'_i) = (x'_i - x_i, y'_i - y_i)$. Ideally, the image-based flow (IBF) vector (u_i, v_i) and the shape-based flow (SBF) vector (u'_i, v'_i) should be the same. A similar analysis can be applied to Camera 2. Hence, we design a novel error metric as follows:

$$EPE(k) = \sqrt{(u_k - u'_k)^2 + (v_k - v'_k)^2}, k \in \Omega_1 \cup \Omega_2, \tag{7}$$

which is based on the popular endpoint error (EPE) used in evaluating optical flow results [4]. Ω_c denotes the pixel set of the cth camera.

It is noteworthy that the proposed error metric Eq.(7) is different from the one used in [19]. Their error metric requires to compute the inverses of the correspondence functions $\mathbf{P}_1()$ and $\mathbf{P}_2()$, which unfortunately may not be generally invertible when multiple pixels receive contributions from the same point. In contrast, our metric does not have such a problem.

In practice, a coarse-to-fine optimization procedure is implemented to search for both the optimal depth map and the best refractive index. We first downsample the acquired correspondence functions $\mathbf{P}_1()$ and $\mathbf{P}_2()$ to $1/4$ of the original resolution. Then, for each index hypothesis in a given range, we optimize Eq.(6) and evaluate the produced depths based on Eq.(7) under the coarse resolution. The index value that gives the smallest reconstruction error is selected. The final shape is reconstructed using the full correspondence functions and the optimal index.

4. Experiments

The proposed approach is evaluated on both synthetic and captured data. The parameter settings $\alpha = \beta = 1, \gamma = 1000, \lambda = 100$ (*unit*) are used in synthetic data and $\alpha = \beta = 1, \gamma = 20, \lambda = 0.005$ (*mm*) are used in real experiments. During the coarse-to-fine minimization, the maximum iteration numbers of L-BFGS-B optimization are fixed to 200 and 20 for the downsampled and full resolutions, respectively, for the first frame. The iteration numbers are reduced by half for the remaining frames. We use the 5×5 and 3×3 local neighborhoods $\mathcal{N}()$ in Eq.(1) at the low and full resolution, respectively. Consider computing the normals $\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_p$ for different points can be performed independently. We implement our algorithm employing parallelization in MATLAB R2016a on an 8-core PC with 3.2GHz Intel Core i7 CPU and 24GB RAM.

4.1. Synthetic Data

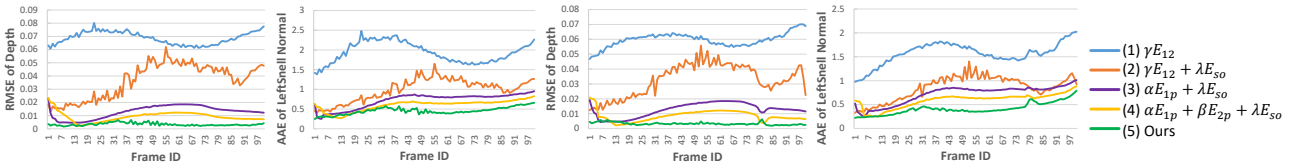
We first validate our approach on a synthetic sinusoidal wave: $z(x, y, t) = 2 + 0.1\cos(\pi(t + 50)\sqrt{(x-1)^2 + (y-0.5)^2}/80)$. In practice, the two cameras are placed at $(0, 0, 0)$ and $(0.05, 0, 0)$, respectively. The reference plane is at $z = 2.5$. By mapping a Bernoulli pattern on the reference plane, we start with rendering the reference image $\mathbf{B}$ without the fluid. Then the distorted image with the wavy surface is simulated using a ray-tracer as illustrated in Fig. 2. The correspondence functions are obtained by performing the correspondence matching algorithm in Sec. 3.1.

The proposed approach is evaluated using the following two measures: the root mean square error (RMSE) between the ground-truth depths and the computed ones, and the average angular error (AAE) between the true normals and the recovered *LeftSnell* normals. Here the *LeftSnell* normals, which can be generated by both the existing method [19] and our approach, are used for fair comparisons. The *PCA* and *RightSnell* normals are used in our formulation only and the evaluation results based on these two normals are similar to the ones presented here; see supplemental materials [1].

To validate the effectiveness of the proposed constraints, we first evaluate the algorithm by removing different terms from Eq.(6). The objective function used in each case is listed in Fig. 3(e). Case 1 includes the cross-view term E_{12} only and corresponds to that used in the previous method [19]. Adding a spatial smoothness term (Case 2) can effectively reduces the errors and hence, the smoothness term is used for all other comparisons with [19]. Case 3 is equivalent to a single-view solution, where only the correspondence information from Camera 1 is used. Case 4 uses E_{1p} and E_{2p}, whereas our approach incorporates all three normal consistency constraints Eq.(2,3,4) in the objective function Eq.(6) and yields the smallest errors. Moreover, Fig. 3 also shows the robustness and temporal coherence of our approach over time.

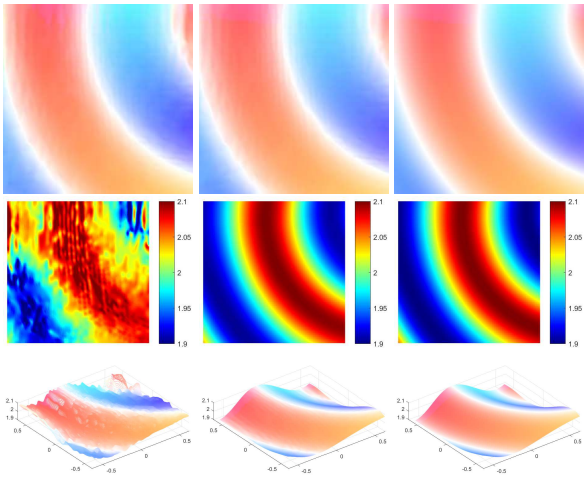
Fig. 4 compares the conventional stereo-based method [19] with ours. For fair comparisons, the pixel-point correspondences generated using our approach are used. The results show that, with added smoothness constraint, their estimated normal maps are similar to ours. However, their estimated depths are noisy whereas ours are smooth. More importantly, our approach simultaneously recovers the depths and the normals, which are both accurate and consistent with each other.

In addition to obtaining the 3D fluid surfaces, our approach can recover the refractive index of the fluid. Here we test the reliability of refractive index estimation. By setting different refractive indices in simulation, we render the distorted images with the fluid using our ray-tracer. As shown in Fig. 5, for each ground-truth index setting, we reconstruct the 3D shape and compute the average EPE Eq.(7) under each index hypothesis in the range of $[1.25, 1.85]$ with increments of 0.05. The EPE curve exhibits a minimum that



(a) Index $\eta_l = 1.33$ (b) Index $\eta_l = 1.33$ (c) Index $\eta_l = 1.55$ (d) Index $\eta_l = 1.55$ (e) Legend

Figure 3. Different error measures as a function of frame id for synthetic wave. (a) and (b) shows the error plots when the refractive index $\eta_l = 1.33$ is used in wave simulation. (c) and (d) shows the error curves when $\eta_l = 1.55$ is used in wave simulation.



(a) [19]+Smoothness (b) Ours (c) Ground Truth

Figure 4. Visual comparisons between the method in [19] and ours for an example frame when $\eta_l = 1.33$ is used for simulation. From top to bottom, it shows the *LeftSnell* normal map, the depth map and the point cloud colored with *LeftSnell* normals. Please see the supplemental materials [1] for the full video sequence as well as the captured images.

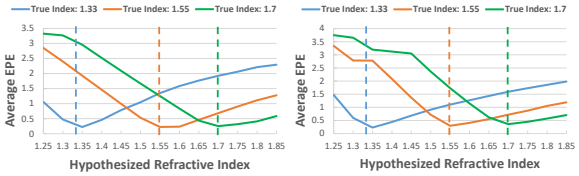


Figure 5. Average EPE Eq.(7) as a function of refractive index hypotheses for synthetic data. The vertical dashed lines indicate the true indices. The left and right figure shows refractive index estimation on the 0th and 45th frame, respectively. Evaluations on other testing frames can be found in the supplemental materials [1].

is close to the true refractive index, which demonstrates that the new error metric Eq.(7) can effectively estimate the refractive index.

4.2. Real Dynamic Water Surfaces

In order to capture real fluid surfaces, we set up a system as shown in Fig. 1(a). Three synchronized Point Grey Flea2 cameras are used for capturing video at 30fps at a resolution of 516×388. Cameras 1 and 2 are used for 3D reconstruction and refractive index estimation, whereas Camera 3 is used *for accuracy assessment only*. We print our binary random patterns on A4-sized papers using a commodity printer. The pattern is then laminated to be waterproof. The refraction effect caused by the thin laminated plastic layer is negligible. The pattern is attached to the bottom of the tank. Another feasible but more expensive solution is to use a waterproof tablet for displaying patterns. Before adding water, we calibrate the relative poses between the cameras and the pattern using a checkerboard [35].

In Fig. 6, three captured water waves are shown and the full sequences can be found in the supplemental videos [1]. Both Wave 1 and Wave 2 are generated by randomly perturbing the water surface at one end of the tank and both exhibit large water fluctuations and fast evolutions. However, two different Bernoulli random patterns with different block sizes are used for evaluating the robustness of the proposed algorithm against pattern changes; see Fig. 7. Wave 3 is a small rippled case generated by dripping water drops near one side of the pattern. Our approach can faithfully recover the propagating annular structures produced by the water drops.

Novel View Synthesis. To evaluate reconstruction quality, we first use the reconstructed surface shape to synthesize the view at Camera 3 and visually compare it with the image observed by the camera. In particular, we first compute the IBF field at Camera 3 using the observed image $\mathbf{I}_3$ and the reference image $\mathbf{B}_3$ as discussed in Sec. 3.1. We then compute the SBF field of Camera 3 from the reconstructed 3D surface using the ray-tracing method as discussed in Sec. 3.3 and shown in Fig. 2. We can now warp $\mathbf{B}_3$ using either the IBF or the SBF to obtain the synthesized image $IBF(\mathbf{B}_3)$ and $SBF(\mathbf{B}_3)$, respectively[2]. By comparing the captured image $\mathbf{I}_3$ with $IBF(\mathbf{B}_3)$ and $SBF(\mathbf{B}_3)$, we can qualitatively evaluate the accuracy of pixel-point correspondences and the quality of 3D reconstruction, respectively.

As shown in Fig. 7, our approach can faithfully synthesize the observations at Camera 3, whereas the results of [19] look quite different. The comparison also shows

[2]Here we use the italic form $IBF()$ and $SBF()$ to denote the functions that compute the synthesized image using IBF and SBF, respectively.

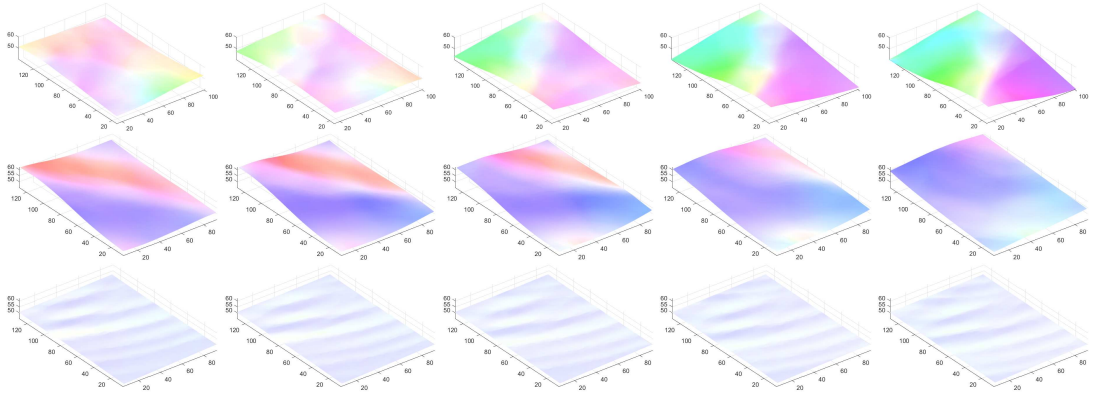
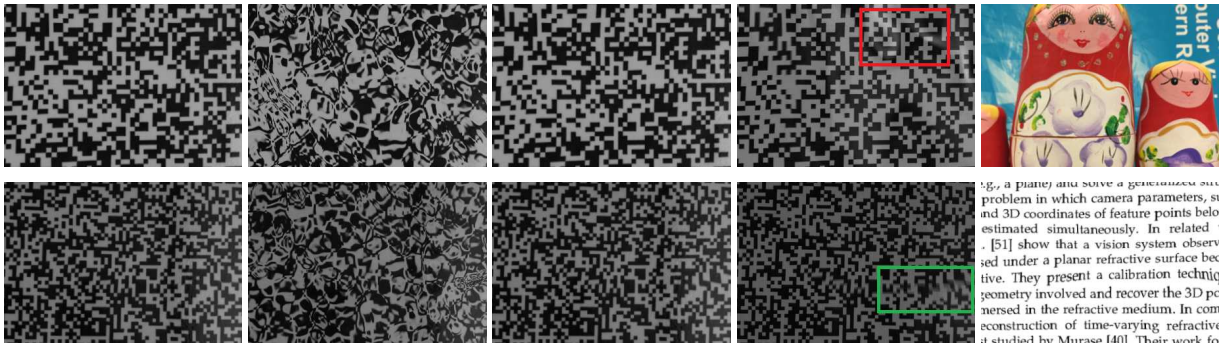
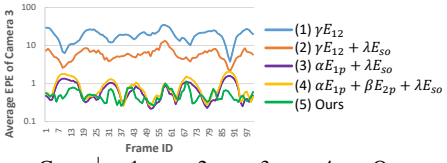


Figure 6. Point clouds of adjacent frames of Wave 1 (top), Wave 2 (middle), Wave 3 (bottom). It shows that our results are visually temporal coherent. In this paper, a point cloud is plotted based on its corresponding depth map and colored with *LeftSnell* normals.



(a) $IBF(\mathbf{B}_3)$ (b) $SBF(\mathbf{B}_3)$ [19] (c) Our $SBF(\mathbf{B}_3)$ (d) Captured Image $\mathbf{I}_3$ (e) Composite

Figure 7. View synthesis using an example frame of Wave 1 (top) and Wave 2 (bottom). The shading effects caused by reflection/caustics (red box) and motion blur effects (green box) can be observed in captured images (d). In (e), we compose the reconstructed 3D surface onto new scenes using the ray-tracing method as depicted in Fig. 2.



Case	1	2	3	4	Ours
Mean	18.90	5.96	0.68	0.86	**0.50**
Stdev.	6.48	2.01	0.36	0.48	**0.17**

Figure 8. Quantitative evaluation on results generated under different constraints using Wave 1. The top figure plots average EPE as a function of frame ID. For better visualization, the 10-base log scale is used for the vertical axis. The bottom table shows the corresponding mean and standard deviation (stdev.) of EPE among all frames.

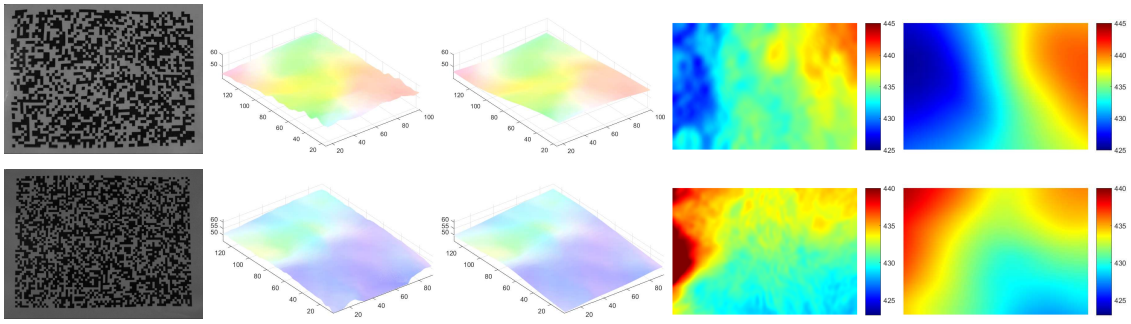
that: 1) the water surface may reflect environment light and may generate caustics, which cause intensity differences between the synthesize view and the captured image; and 2) the water surface moves very fast, which causes motion blur in captured images and is not generated in synthesized view.

Effectiveness of Constraints. Our next experiment aims to quantitatively verify whether or not the novel single-view consistency constraints can help to improve reconstruction accuracy on real data. Since ground truth surfaces are difficult to obtain for real waves, we here use the EPE measure Eq.(7) between the IBF and SBF computed at Camera 3 as explained above. If the IBF is properly estimated and the surface shape is accurately reconstructed, the two flow fields should be consistent. Note that here we do not compare intensity difference between $\mathbf{I}_3$ and $SBF(\mathbf{B}_3)$ because we want to ignore shading differences discussed above and properly evaluate the surface reconstruction error[3].

As shown in Fig. 8, the presented approach achieves the smallest average EPE, which suggests that the 3D shape reconstructed from two views (Camera 1 and 2) is the most consistent with the pixel-point correspondences acquired from the additional view (Camera 3).

Comparisons with [19]. Fig. 9 visually compares our approach and the traditional method [19] on our real waves. Because of the global formulation, our depths and normals are both consistent with the observed image distortions. Our

[3]We also provide additional evaluations by comparing the binarized $\mathbf{I}_3$ and $SBF(\mathbf{B}_3)$ in the supplemental materials [1].



(a) Captured Image $\mathbf{I}_1$ (b) Point Cloud [19] (c) Our Point Cloud (d) Depth Map [19] (e) Our Depth Map

Figure 9. Visual comparisons between the method of [19] and ours for an example frame of Wave 1 (top) and Wave 2 (bottom). Note that here we also impose a smoothness term in the algorithm of [19], *i.e*. Case 2, for fair comparisons.

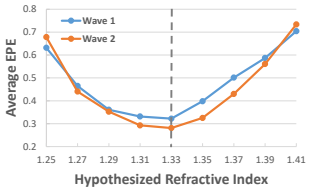


Figure 10. Average EPE Eq.(7) as a function of refractive index hypotheses for real data. The vertical dashed line indicate the refractive index of water, *i.e*. 1.33.

normals also reconcile with the obtained point clouds. In comparison, their normal map looks similar to ours but their depth map is noisy, which is consistent with the reported results in their paper.

Refractive Index Estimation. Following the previous work [19], we compute the average EPE Eq.(7) among 10 frames under different hypothesized refractive indices in the range of $[1.25, 1.41]$ with increments of 0.02. The minima of both curves in Fig. 10 are each close to the refractive index of water, *i.e*. 1.33.

5. Conclusions

We revisit the problem of dynamic refraction stereo [19] by presenting a new global optimization-based framework. We first formulate an objective function which couples both the conventional cross-view normal consistency constraint and the new single-view normal consistency priors that take local surface geometry into consideration. By solving all surface points at the same time, we obtain accurate and consistent depths and normals. Most importantly, our approach successfully avoids the fundamental limitation of previous methods that require using surface integration without accurate boundary conditions. Furthermore, we develop a novel error metric which can reliably estimate the refractive index of liquid in a computer vision fashion. It is also noteworthy that our reconstructed fluid surfaces are highly accurate for the application of novel view synthesis, which cannot be achieved in existing methods.

Our approach works under several common assumptions as in previous refraction-based methods: (i) the fluid is homogeneous and clean, through which light is refracted exactly once, (ii) the pixel-point correspondences can be accurately acquired and (iii) the fluid waves are sufficiently smooth. However, real-world fluid phenomena, which include *e.g*. bubbles, scattering, breaking waves, are created by bending light more than once and thus can violate the above assumptions. In addition, grown out of the surface smoothness assumption, we also assume that the normal at a 3D point can be reliably estimated with its neighboring samples by PCA.

We plan to improve our approach in the following directions. First, the optical flow-based correspondence matching algorithm cannot handle pattern elimination/separation [20] caused by severe distortions. In the future, we plan to investigate two alternative solutions: the single-shot pattern coding [9] techniques and the temporal environment matting methods [6, 23] with high-speed acquisition rates. Second, we will identify conditions under which a unique solution exists and beyond which the type of ambiguous surfaces that may result [19]. Third, we are also interested in removing the flat background constraint by recovering fluid surfaces in natural scenes [33].

Acknowledgments. We thank NSERC, AITF and the University of Alberta for the financial support. Constructive comments from anonymous reviewers and the area chair are highly appreciated. We thank Mr. Jian Wang for useful discussions and Mr. Steve Sutphen, Mr. Li He for experimental assistances.

References

[1] Supplemental materials. http://webdocs.cs.ualberta.ca/~yang/index.html. 5, 6, 7

[2] R. J. Adrian. Particle-imaging techniques for experimental fluid mechanics. *Annual review of fluid mechanics*, 23(1):261–304, 1991. 1

[3] 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 *ACM Transactions on Graphics (TOG)*, page 132. ACM, 2008. 2

[4] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92(1):1–31, 2011. 5

[5] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *European Conference on Computer Vision*, pages 25–36. Springer, 2004. 3, 4

[6] 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 *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 121–130. ACM Press/Addison-Wesley Publishing Co., 2000. 3, 8

[7] Y. Ding, F. Li, Y. Ji, and J. Yu. Dynamic fluid surface acquisition using a camera array. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2478–2485. IEEE, 2011. 1, 2, 3, 4

[8] D. Enright, S. Marschner, and R. Fedkiw. Animation and rendering of complex water surfaces. In *ACM Transactions on Graphics (TOG)*, volume 21, pages 736–744. ACM, 2002. 1

[9] J. Geng. Structured-light 3d surface imaging: a tutorial. *Advances in Optics and Photonics*, 3(2):128–160, 2011. 3, 8

[10] J. Gregson, I. Ihrke, N. Thuerey, and W. Heidrich. From capture to simulation: connecting forward and inverse problems in fluids. *ACM Transactions on Graphics (TOG)*, 33(4):139, 2014. 1

[11] K. Han, K.-Y. K. Wong, and M. Liu. A fixed viewpoint approach for dense reconstruction of transparent objects. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4001–4008, 2015. 2

[12] I. Ihrke, K. N. Kutulakos, H. Lensch, M. Magnor, and W. Heidrich. Transparent and specular object reconstruction. In *Computer Graphics Forum*, volume 29, pages 2400–2426. Wiley Online Library, 2010. 2

[13] 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 STAR–STATE OF THE ART REPORT*. Citeseer, 2008. 2

[14] B. Jähne, J. Klinke, and S. Waas. Imaging of short ocean wind waves: a critical theoretical review. *JOSA A*, 11(8):2197–2209, 1994. 1, 2

[15] Y. Ji, J. Ye, and J. Yu. Reconstructing gas flows using light-path approximation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2507–2514, 2013. 2

[16] K. N. Kutulakos and E. Steger. A theory of refractive and specular 3d shape by light-path triangulation. *International Journal of Computer Vision*, pages 13–29, 2008. 2

[17] C. Li, D. Pickup, T. Saunders, D. Cosker, D. Marshall, P. Hall, and P. Willis. Water surface modeling from a single viewpoint video. *IEEE Transactions on Visualization and Computer Graphics*, 19(7):1242–1251, 2013. 2

[18] M. Liu, R. Hartley, and M. Salzmann. Mirror surface reconstruction from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(4):760–773, 2015. 2

[19] N. J. Morris and K. N. Kutulakos. Dynamic refraction stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1518–1531, 2011. 1, 2, 3, 4, 5, 6, 7, 8

[20] H. Murase. Surface shape reconstruction of a nonrigid transparent object using refraction and motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(10):1045–1052, 1992. 2, 8

[21] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C*, volume 2. Cambridge university press Cambridge, 1996. 4

[22] Y. Qian, M. Gong, and Y. Hong Yang. 3d reconstruction of transparent objects with position-normal consistency. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4369–4377, 2016. 2, 4

[23] Y. Qian, M. Gong, and Y.-H. Yang. Frequency-based environment matting by compressive sensing. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3532–3540, 2015. 8

[24] R. B. Rusu. *Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments*. PhD thesis, Computer Science department, Technische Universitaet Muenchen, Germany, October 2009. 4

[25] Q. Shan, S. Agarwal, and B. Curless. Refractive height fields from single and multiple images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 286–293. IEEE, 2012. 2, 4

[26] K. Tanaka, Y. Mukaigawa, H. Kubo, Y. Matsushita, and Y. Yagi. Recovering transparent shape from time-of-flight distortion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4387–4395, 2016. 2

[27] J. V. Uspensky. Introduction to mathematical probability. Technical report, 1937. 3

[28] G. Wetzstein, R. Raskar, and W. Heidrich. Hand-held schlieren photography with light field probes. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–8. IEEE, 2011. 2

[29] G. Wetzstein, D. Roodnick, W. Heidrich, and R. Raskar. Refractive shape from light field distortion. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1180–1186. IEEE, 2011. 1, 2, 3

[30] T. Xue, M. Rubinstein, N. Wadhwa, A. Levin, F. Durand, and W. T. Freeman. Refraction wiggles for measuring fluid depth and velocity from video. In *European Conference on Computer Vision*, pages 767–782. Springer, 2014. 2

[31] J. Ye, Y. Ji, F. Li, and J. Yu. Angular domain reconstruction of dynamic 3d fluid surfaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 310–317. IEEE, 2012. 1, 2

[32] M. Yu and H. Quan. Fluid surface reconstruction based on specular reflection model. *Computer Animation and Virtual Worlds*, 24(5):497–510, 2013. 2

[33] M. Zhang, X. Lin, M. Gupta, J. Suo, and Q. Dai. Recovering scene geometry under wavy fluid via distortion and defocus analysis. In *European Conference on Computer Vision*, pages 234–250. Springer, 2014. 8

[34] X. Zhang and C. S. Cox. Measuring the two-dimensional

structure of a wavy water surface optically: A surface gradient detector. *Experiments in Fluids*, 17(4):225–237, 1994. 1, 2

[35] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11):1330–1334, 2000. 6

[36] 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 (TOMS)*, 23(4):550–560, 1997. 4

[37] X. Zuo, C. Du, S. Wang, J. Zheng, and R. Yang. Interactive visual hull refinement for specular and transparent object surface reconstruction. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2237–2245, 2015. 2