Recovering Scene Geometry under Wavy Fluid via Distortion and Defocus Analysis*

Mingjie Zhang[1], Xing Lin[1], Mohit Gupta[2], Jinli Suo[1], and Qionghai Dai[1]

[1] Department of Automation, Tsinghua University
[2] Columbia University

Abstract. In this paper, we consider scenes that are immersed in transparent refractive media with a dynamic surface. We take the first steps to reconstruct both the 3D fluid surface shape and the 3D structure of immersed scene simultaneously by utilizing distortion and defocus clues. We demonstrate that the images captured through a refractive dynamic fluid surface are the distorted and blurred versions of all-in-focused (AIF) images captured through a flat fluid surface. The amounts of distortion and refractive blur are formulated by the shape of fluid surface, scene depth and camera parameters, based on our refractive geometry model of a finite aperture imaging system. An iterative optimization algorithm is proposed to reconstruct the distortion and immersed scene depth, which are then used to infer the 3D fluid surface. We validate and demonstrate the effectiveness of our approach on a variety of synthetic and real scenes under different fluid surfaces.

Keywords: underwater 3D reconstruction, dynamic fluid surface recovery, refractive blur, distortion, depth from defocus.

1 Introduction

In recent years, the problems of recovering scene structure immersed in refractive fluid and reconstructing the 3D shape of the dynamic fluid surface have drawn more attention in multiple research fields, including computer vision and oceanography. Although a lot of progress has been made [30,9,36,8,11,4], the solutions are not sufficiently general to be used in real-world scenarios. This is because most existing methods recovering scene geometry under fluid surface assume the fluid surface to be flat [8,34,3]. On the other hand, 3D fluid surface estimation approaches [22,9,24,30] assume a flat scene under the surface. These assumptions are rarely satisfied in real applications. The general scenario of recovering scene depths immersed in fluid with non-flat surfaces remains a challenging problem. In addition, most previous approaches are based on the pinhole imaging model. However, in practice, in order to achieve high signal-to-noise-ratio (SNR), cameras often use large apertures where pinhole model is not applicable, thus resulting in image blur.

* Electronic supplementary material - Supplementary material is available in the online version of this chapter at `http://dx.doi.org/10.1007/978-3-319-10602-1_16`. Videos can also be accessed at `http://www.springerimages.com/videos/978-3-319-10601-4`

D. Fleet et al. (Eds.): ECCV 2014, Part V, LNCS 8693, pp. 234–250, 2014.

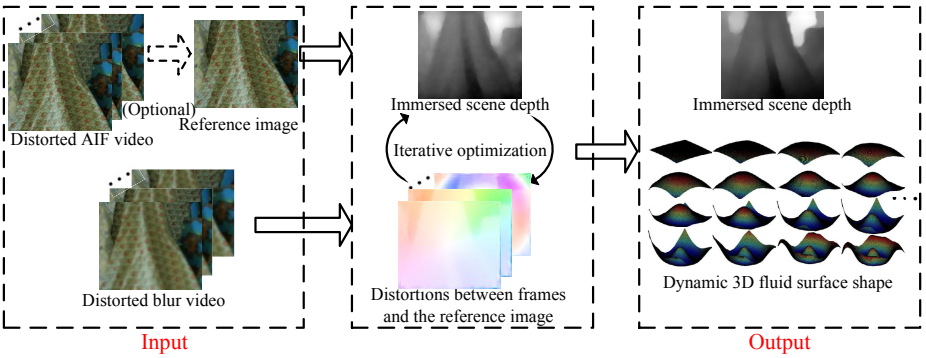


Fig. 1. Diagram of the proposed iterative optimization framework. The input of our approach are a captured refractive blur video and a reference image which could be estimated from another all-in-focused(AIF) video or captured under flat water. An iterative optimization is applied to recover the depth of immersed scene and the distortions, which are then used to estimate the dynamic 3D fluid surface shape.

Among the large numbers of depth estimation approaches (e.g. multi-view stereo, structure from motion, shape from shading), depth from defocus (DFD) is attractive due to its insensitivity to an occlusion and matching problem [29]. Different from performing DFD in clear air, the irregular refraction on the wavy interface causes distortion and blurring of the images of immersed scenes. The blur in images captured through a fluid surface is determined by not only camera parameters and scene depth but also the refraction on the fluid surface. Hence, we call it the **refractive blur.** Compared with stereo, DFD approach has a smaller baseline. Thus, all the rays emitted from a scene point reaching the sensor can be assumed to be sufficiently close so that the normals of the fluid surface where the rays cross the fluid interface can be assumed to be approximately constant. This reduces the number of unknowns as compared to stereo, where rays from a scene point cross the fluid interface at different points, and thus likely encounter different surface normals (details in the Sec. 6).

In this paper, we establish a geometric imaging model for refractive blur and distortion simultaneously, while most existing works do not account for refractive blur. Our imaging model represents the images captured through the fluid surface as the distorted and blurred version of the undistorted all-in-focus(AIF) image captured through a flat fluid surface.

The reconstruction steps of fluid surface and the underneath scene geometry are illustrated in Fig. 1. Our algorithm requires an out-of-focus video captured under large aperture setting and a reference image[1]. The reference image could be estimated from the pre-captured AIF video or captured under flat water with a small aperture. Then, based on the model established in Sec. 3, we construct an objective function and use an optimization procedure to compute the depth of the immersed scene and the distortions alternatively. Finally the dynamic 3D

[1] In this paper, the reference image refers to the undistorted all-in-focus(AIF) image captured under flat water with a small aperture.

fluid surface shape video is also recovered from the depth and distortions maps based on the established geometry model.

Specifically, the paper has the following contributions:

- We establish the refractive blur and distortion geometry model as a function of the camera settings, the shape of fluid surface and scene structure.
- We present a novel iterative global optimization method for recovering under-fluid scene structure and 3D fluid surface from distortion and refractive blur.
- We obtain promising results on both synthetic and real captured data.

2 Related Work

Fluid Surface Reconstruction. Several methods [22,9,24,30] recover the wavy fluid surface and undistorted image by analyzing the distortion in the video. These works require placing a flat plane with rich features under the fluid surface. In order to enhance the reconstruction performance, multi-camera methods [22,9,18] have also been proposed. Moreover, Tian and Narasimhan [30] model the distortion by the wave equation and present a tracking method without the need of undistorted image. There are also methods utilizing active illumination instead of a flat rich feature plane, such as [19,36]. Our paper estimates scene structure as well as fluid appearance, does not require active illumination, and owns wider applications.

Compensation of the Refractive Distortions. A variety of methods [11,10,33,31,26] have been proposed for removing the non-rigid distortions in the captured images without recovering the shape of the fluid surface. Most of these approaches adopt the lucky imaging strategy by seamlessly stitching the patches with least distortions, which can be searched or calculated via various techniques, such as clustering [11,10], iterative averaging [26], bispectral analysis [33] and progressive warping [31]. These works can be used for providing the reference image from a captured AIF sequence as an input for our algorithm.

Reconstruction of Geometry through Fluid. Reconstruction of 3D structure under or above the water surface is also an active area of research [8,4,3,14]. Chang and Chen [8] and Ferreira et al. [14] apply the structure from motion and stereo methods to reconstruct the 3D structure of scenes submerged in refractive fluid, respectively. A stochastic triangulation method is proposed by Alterman et al. [4] to recover the structure of scene above water from a video pair captured under water. These works provide some preliminary studies but are limited to static and flat fluid surface.

Depth from Defocus (DFD) in Clear Medium. DFD approaches capture two or multiple defocused images under different focal settings for recovering the scene structure [13,12,21]. These approaches assume that both the scene and the camera are in the same and clear medium. Applying DFD where the scene and the camera are immersed in different refractive media has received little attention. In this paper, we establish a refractive blur model to generalize the conventional defocus model and exploit it to estimate the scene depths under dynamic fluid surface using a reference image and a refractive blur video.

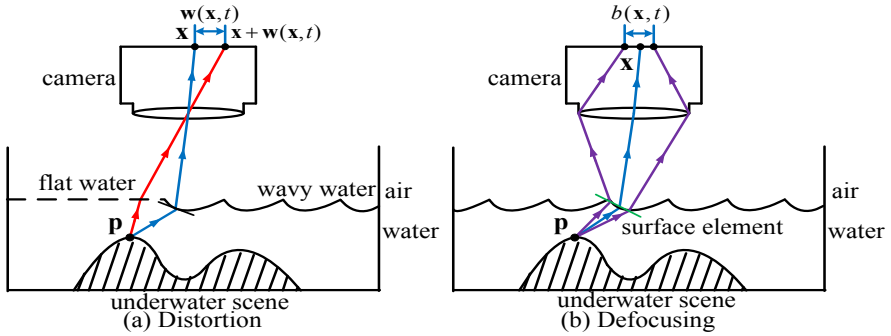


Fig. 2. Imaging through fluid surface with distortion (a) and defocusing (b). (a) The pixel $\mathbf{x} + \mathbf{w}(\mathbf{x}, t)$ in the reference image $I(\mathbf{x}, 0)$ is refracted to $\mathbf{x}$ at frame t. (b) The refracted light rays are further blurred into a refractive blur size $b(\mathbf{x}, t)$ due to the finite camera aperture size.

Other related works to ours include surface reconstruction of transparent refraction object [34,5,23,17], and camera calibration for imaging through the water-air interface [32,35].

3 The Refractive Blur and Distortion Geometry Model

3.1 Image Formulation Model

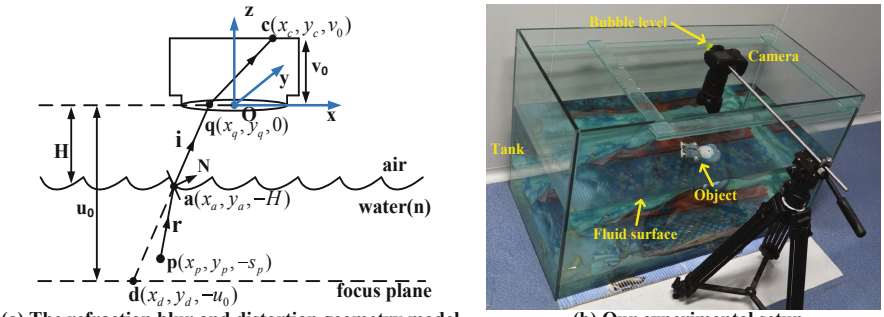
As shown in Fig. 2, this work supposes that a static non-plane scene is placed under a dynamic fluid surface, and a video camera is focused on a certain plane. For simplicity, we ignore scattering, light absorption and chromatic dispersion in the fluid. We assume that the exposure time is short enough to ignore the motion blur. Let $I(\mathbf{x}, t)$ denotes the t^{th} AIF video frame taken through the fluid surface with $\mathbf{x} = \{x, y\}$ being the 2D spatial coordinates, and $I(\mathbf{x}, 0)$ denotes the reference image.

Without refractive blur, each AIF video frame $I(\mathbf{x}, t)$ is a distorted version of $I(\mathbf{x}, 0)$ as shown in Fig. 2 (a) and can be expressed as

$$I(\mathbf{x}, t) = I(x, y, t) = I(\mathbf{x} + \mathbf{w}(\mathbf{x}, t), 0) = I(x + u(\mathbf{x}, t), y + v(\mathbf{x}, t), 0), \quad (1)$$

where $\mathbf{w}(\mathbf{x}, t) = (u(\mathbf{x}, t), v(\mathbf{x}, t))$ denotes the distortion between the reference image and frame i correspondence to the point $\mathbf{x}$ at frame t, which will be formulated in Sec. 3.2 in detail.

Then considering both the refraction and the defocusing, as shown in Fig. 2 (b), rays emitted from an underwater scene point are deflected at the water surface and projected onto different positions of the sensor. Similar to the defocus blur occurring in the air, we call this deflection blur as "refractive blur". Detailed formulation of refractive blur will also be given in Sec. 3.2. Thus the refractive blur video acquired through a wavy water surface can be formulated as



(a) The refraction blur and distortion geometry model (b) Our experimental setup

Fig. 3. The refractive geometry of underwater imaging in 3D case (a), and our experiment setup (b). (a) One ray emitted from the scene point **p** is refracted at the fluid surface point **a** , passes through the aperture plane at **q** and is projected to point **c** on the sensor, which corresponds to point **d** on the focus plane. The coordinates axes are displayed with blue arrows and the origin is set at the camera optical center **O**. (b). Our camera's sensor is parallel with the flat water surface which is achieved by using bubble level.

$$B(\mathbf{x},t)=\int_{\mathbf{y}\in N_{\mathbf{x}}} h_{\sigma^2(\mathbf{x},t)}(\mathbf{x},\mathbf{y})I(\mathbf{y},t)d\mathbf{y}=\int_{\mathbf{y}\in N_{\mathbf{x}}} h_{\sigma^2(\mathbf{x},t)}(\mathbf{x},\mathbf{y})I(\mathbf{y}+\mathbf{w}(\mathbf{y},t),0)d\mathbf{y}, \quad (2)$$

where $B(\mathbf{x},t)$ denotes the t^{th} frame in the refractive blur video; $h_{\sigma^2(\mathbf{x},t)}(\mathbf{x},\mathbf{y})$ denotes the refractive blur kernel of point $\mathbf{x}$ which can be approximately estimated by Gaussian kernel; $\sigma(\mathbf{x},t)$ denotes the kernel size of point $\mathbf{x}$ at time t; and $\mathbf{y}\in N_{\mathbf{x}}$ denotes the pixels within $\mathbf{x}$'s neighborhood.

3.2 The Refractive Blur and Distortion Formulation

As shown in Fig. 3, suppose that the sensor is parallel with the flat water surface. H denotes the vertical distance from the camera to the flat water surface, n denotes the refractive index of water, u_0 denotes the distance between the lens and the focus plane, v_0 denotes the distance between the lens and the sensor plane, and the focal plane remains unchanged during the capture. Then based on the vector forms of Snell's law [20] in 3D space, we can obtain the relationship

$$\boldsymbol{i}\times\boldsymbol{N}=n\boldsymbol{r}\times\boldsymbol{N}, \quad (3)$$

where $\boldsymbol{i}$ denotes the unit vector along the light propagating in the air; $\boldsymbol{r}$ denotes the unit vector along the corresponding light propagating in the water; $\boldsymbol{N}$ denotes the unit normal vector of the refractive plane at the intersection point of $\boldsymbol{i}$ and $\boldsymbol{r}$; and $\times$ is the cross product.

In general 3D case, we use the thin lens model to analyze the light path and assume the x-axis and y-axis are parallel to the sensor plane, z-axis is aligned with the camera optical center **o**, represented by blue arrows in Fig. 3(a). Without the loss of generality, we also assume the amplitude of water wave is slight enough

to ignore. Then Eq. 3 can be represented by the geometric locations of the underwater scene point $\mathbf{p} = (x_p, y_p, -s_p)$, the refractive point $\mathbf{a} = (x_a, y_a, -H)$ on the refractive plane and the intersection point $\mathbf{q} = (x_q, y_q, 0)$ on the aperture plane

$$\boldsymbol{i} = \frac{\mathbf{q}-\mathbf{a}}{\|\mathbf{q}-\mathbf{a}\|}\ ,\ \boldsymbol{r} = \frac{\mathbf{a}-\mathbf{p}}{\|\mathbf{a}-\mathbf{p}\|}. \tag{4}$$

As shown in Fig.3(a), this ray emitted from the immersed scene point $\mathbf{p}$ is projected to the sensor point $\mathbf{c}$, which corresponds to the point $\mathbf{d}$ on the focus plane. Based on the simple lens model, these two points have the relationship: $\mathbf{d} = (u_0/v_0)\mathbf{c}$. The three points $\mathbf{q}$, $\mathbf{a}$ and $\mathbf{d}$ are also collinear, thus point $\mathbf{a}$ can be represented by

$$\mathbf{a} = (1-\frac{H}{u_0})\mathbf{q} - \frac{H}{u_0}\mathbf{d} = (1-\frac{H}{u_0})\mathbf{q} - \frac{H}{v_0}\mathbf{c}. \tag{5}$$

Substitute Eq. 4 and Eq. 5 into Eq. 3, we can obtain the geometric function represented one ray emitted from the scene point $\mathbf{p}$ and projected to the sensor point $\mathbf{c}$, which also provides the relationship between these two points:

$$\begin{cases} \frac{(x_q+\frac{u_0}{v_0}x_c)n_z(\mathbf{a})-u_0 n_x(\mathbf{a})}{\sqrt{\left(x_q+\frac{u_0}{v_0}x_c\right)^2+\left(y_q+\frac{u_0}{v_0}y_c\right)^2+{u_0}^2}} = n\frac{((1-\frac{H}{u_0})x_q-\frac{H}{v_0}x_c-x_p)n_z(\mathbf{a})-(s_p-H)n_x(\mathbf{a})}{\sqrt{\left((1-\frac{H}{u_0})x_q-\frac{H}{v_0}x_c-x_p\right)^2+\left((1-\frac{H}{u_0})y_q-\frac{H}{v_0}y_c-y_p\right)^2+(s_p-H)^2}} \\ \frac{(y_q+\frac{u_0}{v_0}y_c)n_z(\mathbf{a})-u_0 n_y(\mathbf{a})}{\sqrt{\left(x_q+\frac{u_0}{v_0}x_c\right)^2+\left(y_q+\frac{u_0}{v_0}y_c\right)^2+{u_0}^2}} = n\frac{((1-\frac{H}{u_0})y_q-\frac{H}{v_0}y_c-y_p)n_z(\mathbf{a})-(s_p-H)n_y(\mathbf{a})}{\sqrt{\left((1-\frac{H}{u_0})x_q-\frac{H}{v_0}x_c-x_p\right)^2+\left((1-\frac{H}{u_0})y_q-\frac{H}{v_0}y_c-y_p\right)^2+(s_p-H)^2}} \end{cases}, \tag{6}$$

where $n_x(\mathbf{a})$, $n_y(\mathbf{a})$ and $n_z(\mathbf{a})$ are the 3D coordinates of the unit normal vector $\boldsymbol{N}$ at the refractive point $\mathbf{a}$, and s_p is the depth of the scene point $\mathbf{p}$.

With Eq. 6, we can project the underwater scene point to the sensor plane or conversely, thus the distortion between different frames can be calculated. By scanning the point $\mathbf{q}$ over the aperture plane while holding the scene point $\mathbf{p}$, we can also derive the blur kernel size of the scene point $\mathbf{p}$ on the sensor. Unfortunately, Eq. 6 is too complex to analyze. Therefore, we use the first order Taylor expansion for simplification (details in the Supplementary Material). We regard x_p and y_p as the independent variable and regard x_c and y_c as unknown variable. The first order Taylor expansion of Eq. 6 at point $x_p = x_{p0} = x_q - \frac{n_x(\mathbf{a})}{n_z(\mathbf{a})}s_p$ and $y_p = y_{p0} = y_q - \frac{n_y(\mathbf{a})}{n_z(\mathbf{a})}s_p$ is

$$\begin{cases} x_c \approx \frac{v_0}{u_0}(\frac{n_x(\mathbf{a})}{n_z(\mathbf{a})}u_0 - x_q) - (x_p - x_q + \frac{n_x(\mathbf{a})}{n_z(\mathbf{a})}s_p)\frac{nv_0}{s_p+(n-1)H} \\ y_c \approx \frac{v_0}{u_0}(\frac{n_y(\mathbf{a})}{n_z(\mathbf{a})}u_0 - y_q) - (y_p - y_q + \frac{n_y(\mathbf{a})}{n_z(\mathbf{a})}s_p)\frac{nv_0}{s_p+(n-1)H}. \end{cases} \tag{7}$$

Eq. 7 is the simple approximation function representing the light that is emitted from the scene point $\mathbf{p}$ and passes through the lens point $\mathbf{q}$ with the sensor projection $\mathbf{c}$. When we scan the aperture point $\mathbf{q}$ within the circle aperture to analyze the size of blur kernel, the refractive point $\mathbf{a}$ on the refractive plane

is also changing, which makes the analysis difficult. However, as illustrated in Fig. 2 (b), when the varying area (i.e. the **surface element** shown in Fig. 2(b)) of the refractive point **a** is small enough (The detailed analysis will be presented in Sec. 3.3), we can ignore the changes of the normal vector $\boldsymbol{N}$. Then we can find that the area on the sensor corresponding to the scene point (i.e. the refractive blur kernel) is approximatively a circle, and **the refractive blur size** of pixel $\mathbf{x}$ at frame t is

$$\sigma(\mathbf{x},t) = \kappa\frac{v_0 D}{2}\left|\frac{n}{s(\mathbf{x})+(n-1)H} - \frac{1}{u_0}\right|, \tag{8}$$

where D is the aperture diameter of the lens, $s(\mathbf{x})$ is the depth map of reference image and κ is the calibration parameter converting world coordinate to image plane. Notice that if there is no fluid, i.e. $n = 1$, refractive blur size (Eq. 8) is degraded to the defocused blur size in the air [21]. In addition, Eq. 8 shows that the refractive blur size is independent of the water surface shape which means we can infer the immersed scene depth from refractive blur.

In order to derive the amount of distortion, we apply the perspective model by keeping an infinite small aperture size in our refractive geometry model. Then based on Eq. 7, we can also derive the 2D spatial coordinates **distortion** between reference image and other frames by back-projecting point to the scene through the wavy surface and forward-projecting it to the sensor through the flat surface

$$\begin{cases} u(\mathbf{x},t) = -\kappa(\dfrac{n_x(\mathbf{x},t)}{n_z(\mathbf{x},t)} - \dfrac{n_x(\mathbf{x}+\mathbf{w}(\mathbf{x},t),0)}{n_z(\mathbf{x}+\mathbf{w}(\mathbf{x},t),0)})(v_0 - \dfrac{nv_0 s(\mathbf{x}+\mathbf{w}(\mathbf{x},t))}{s(\mathbf{x}+\mathbf{w}(\mathbf{x},t))+(n-1)H}) \\ v(\mathbf{x},t) = -\kappa(\dfrac{n_y(\mathbf{x},t)}{n_z(\mathbf{x},t)} - \dfrac{n_y(\mathbf{x}+\mathbf{w}(\mathbf{x},t),0)}{n_z(\mathbf{x}+\mathbf{w}(\mathbf{x},t),0)})(v_0 - \dfrac{nv_0 s(\mathbf{x}+\mathbf{w}(\mathbf{x},t))}{s(\mathbf{x}+\mathbf{w}(\mathbf{x},t))+(n-1)H}), \end{cases} \tag{9}$$

where $n_x(\mathbf{x},t)$, $n_y(\mathbf{x},t)$ and $n_z(\mathbf{x},t)$ are the 3D coordinates of the unit normal vector $\boldsymbol{N}$ corresponding to the point(x,y) at frame t.

3.3 The Condition of Model

As illustrated in Fig. 2(b) and the derivation in Sec. 3.2, due to the camera's finite aperture, a scene point **p** emits a cluster of rays projecting to the sensor. Each of these rays is deflected at different points on the water-air interface, and we call the area on the interface where these rays pass through the **surface element** corresponding to point **p**.

When we obtain Eq.8 from Eq.7, we assume that each scene point's surface element is small enough to be approximated by a plane. Thus **the size of surface element** is the minimum area of water surface that our model can distinguish, which is similar to the spatial resolution of traditional camera.

Based on Eq. 5 and Eq. 7, we can derive the coordinates of the refractive point **a**:

$$\mathbf{x}_a = \frac{s_p - H}{s_p+(n-1)H}(x_q + (n-1)H\frac{n_{\mathbf{x}}(\mathbf{x}_a)}{n_z(\mathbf{x}_a)}) + \frac{nHx_p}{s_p+(n-1)H}. \tag{10}$$

Randomly select two points ($\mathbf{a_1}$ and $\mathbf{a_2}$) on the surface element corresponding to the same scene point $\mathbf{p}$, then the distance between them is

$$\mathbf{x}_{a_1} - \mathbf{x}_{a_2} = \frac{s_p - H}{s_p + (n-1)H}((x_{q_1} - x_{q_2}) + (n-1)H(\frac{n_{\mathbf{x}}(\mathbf{x}_{a_1})}{n_z(\mathbf{x}_{a_1})} - \frac{n_{\mathbf{x}}(\mathbf{x}_{a_2})}{n_z(\mathbf{x}_{a_2})})). \quad (11)$$

Assuming the model holds, we can derive the size of surface elements $\triangle_p$ corresponding to scene point $\mathbf{p}$:

$$\triangle_p = \max(\mathbf{x}_{a_1} - \mathbf{x}_{a_2}) = \frac{s_p - H}{s_p + (n-1)H} D \leq D \quad (12)$$

From Eq. 12, we know that the size of surface element is positively correlated with the aperture size, i.e. the size of baseline. Thus the smaller the baseline, the better performance and system's robustness. Obviously the DFD approaches usually have a smaller baseline than stereo method and are more suitable to this model in the paper.

4 Iterative Optimization Algorithm for Reconstructing Scene Depth and Water Surface Shape

In this section, we propose an iterative optimization method to reconstruct both the undistorted scene depth and the wavy water surface.

4.1 Optimization Model

Assume the depth of water, the refractive index of water, and the distance from the camera to the water surface are measured in advance, the camera's sensor is parallel with the water surface, and the focal plane remains unchanged during the capture. Then according to Eq. 2, the optimization problem for estimating the distortion $\mathbf{w}(\mathbf{x}, t)$ and the depth $s(\mathbf{x})$ can be formulated as

$$\min_{s,\mathbf{w}} J(s(\mathbf{x}), \mathbf{w}(\mathbf{x}, t)) = \min_{s,\mathbf{w}} E_d(s(\mathbf{x}), \mathbf{w}(\mathbf{x}, t)) + \alpha E_m(\mathbf{w}(\mathbf{x}, t)) + \beta E_m(s(\mathbf{x})), \quad (13)$$

where $J(s(\mathbf{x}), \mathbf{w}(\mathbf{x}, t))$, $E_d(s(\mathbf{x}), \mathbf{w}(\mathbf{x}, t))$, $E_m(s(\mathbf{x}))$ and $E_m(\mathbf{w}(\mathbf{x}, t))$ are objective function, data term, depth regularization term and distortion regularization term, respectively; $\alpha > 0$ and $\beta > 0$ are the regularization parameters that balance data term and two regularization terms, respectively; $s(\mathbf{x})$ denotes the depth corresponding to reference image $I(\mathbf{x}, 0)$.

Specifically, the data term can be written as

$$E_d(s(\mathbf{x}), \mathbf{w}(\mathbf{x}, t)) = \sum_{t=1}^{T} \int_{\Omega} \psi(\left\| B(\mathbf{x}, t) - \int_{\mathbf{y} \in N_{\mathbf{x}}} h_{\sigma^2(\mathbf{x},t)}(\mathbf{x}, \mathbf{y}) I(\mathbf{y} + \mathbf{w}(\mathbf{y}, t), 0) d\mathbf{y} \right\|_2^2) d\mathbf{x}, \quad (14)$$

and depth and distortion regularization terms are

$$E_m(\mathbf{w}(\mathbf{x},t)) = \sum_{t=1}^{T} \int_{\Omega} \psi(\|\nabla u(\mathbf{x},t)\|_2^2 + \|\nabla v(\mathbf{x},t)\|_2^2) d\mathbf{x}, \tag{15}$$

$$E_m(s(\mathbf{x})) = \int_{\Omega} \|\nabla s(\mathbf{x})\|_2 d\mathbf{x}, \tag{16}$$

where $\Omega \subset \mathbb{R}^2$ is the range of $\mathbf{x}$; $\psi(\xi^2) = \sqrt{\xi^2 + \epsilon^2}$ is applied to reduce the outliers, similar to optical flow algorithms in [7,6]; and the distorted displacement $u(\mathbf{x},t)$, $v(\mathbf{x},t)$ and kernel size $\sigma(\mathbf{x},t)$ have been formulated in Eq. 8 and Eq. 9.

The optimization for Eq. 13 requires that the reference image $I(\mathbf{x},\mathbf{0})$ is known. According to Eq. 9, if we want to recover the normal vectors of dynamic fluid surface from distortions and scene depth, the reference image must be the AIF undistorted image taken through the flat water. In this paper, we capture the reference image directly by using the small aperture size under flat water. Besides, we also estimate the reference image from a captured AIF distorted video by an existing method [26], of which synthetic results are shown in the supplementary material and video.

To optimize Eq. 13, we apply an alternative minimization approach to iteratively estimate the distortion $\mathbf{w}(\mathbf{x},t)$ and blur size $s(\mathbf{x})$ which are detailed in Sec. 4.2 and Sec. 4.3, respectively. Next, surface shape could be reconstructed from $s(\mathbf{x})$ and $\mathbf{w}(\mathbf{x},t)$ as described in Sec. 4.4.

4.2 Distortion Refinement

For the minimization of distortion $\mathbf{w}(\mathbf{x},t)$, we keep the depth $s(\mathbf{x})$ fixed, then the optimization in Eq. 13 can be simplified as

$$\min_{\mathbf{w}} J_1(\mathbf{w}(\mathbf{x},t)) = \min_{\mathbf{w}} E_d(\mathbf{w}(\mathbf{x},t)) + \alpha E_m(\mathbf{w}(\mathbf{x},t)), \tag{17}$$

where $J_1(\mathbf{w}(\mathbf{x},t))$ is the objective function for distortion refinement; $E_d(\mathbf{w}(\mathbf{x},t))$ is the data-term defined in Eq. 14 with fixed $s(\mathbf{x})$; $E_m(\mathbf{w}(\mathbf{x},t))$ is the distortion regularization term defined in Eq. 15.

The distortion refinement objective function in Eq. 17 is similar to but different from the objective function of optical flow algorithm in [7,6] in two aspects: the distorted images $I(\mathbf{x},t)$ in [7] are replaced with the distorted and blur images $\int_{\mathbf{y}\in N_{\mathbf{x}}} h_{\sigma^2(\mathbf{x},t)}(\mathbf{x},\mathbf{y}) I(\mathbf{y}+\mathbf{w}(\mathbf{y},i),0) d\mathbf{y}$; the weight γ of gradient image in [7] is set to zero. So we modify the numerical solution of existing optical flow algorithms in [7,6] for our problem:

$$\begin{cases} \psi'((I_z^k + I_x^k du^{k,l} + I_y^k dv^{k,l})^2) \cdot (I_x^k (I_z^k + I_x^k du^{k,l+1} + I_y^k dv^{k,l+1})) \\ \quad - \alpha div(\psi'(\left|\nabla(u+du^{k,l})\right|^2 + \left|\nabla(v+dv^{k,l})\right|^2) \nabla(u+du^{k,l+1})) = 0, \\ \psi'((I_z^k + I_x^k du^{k,l} + I_y^k dv^{k,l})^2) \cdot (I_y^k (I_z^k + I_x^k du^{k,l+1} + I_y^k dv^{k,l+1})) \\ \quad - \alpha div(\psi'(\left|\nabla(u+du^{k,l})\right|^2 + \left|\nabla(v+dv^{k,l})\right|^2) \nabla(v+dv^{k,l+1})) = 0, \end{cases} \tag{18}$$

where

$$\begin{aligned}
I_z^k &= \int_{\mathbf{y}\in N_{\mathbf{x}}} h_{\sigma^2(\mathbf{x},t)}(\mathbf{x},\mathbf{y}) I(\mathbf{y}+\mathbf{w}(\mathbf{y},i),0) d\mathbf{y} - B(\mathbf{x},i), \\
I_x^k &= \int_{\mathbf{y}\in N_{\mathbf{x}}} h_{\sigma^2(\mathbf{x},t)}(\mathbf{x},\mathbf{y}) \partial_x I(\mathbf{y}+\mathbf{w}(\mathbf{y},i),0) + \frac{\partial h_{\sigma^2(\mathbf{x},t)}(\mathbf{x},\mathbf{y})}{\partial s(\mathbf{x}+\mathbf{w}(\mathbf{x},i))} \partial_x s(\mathbf{x}+\mathbf{w}(\mathbf{x},i)) I(\mathbf{x}+\mathbf{w}(\mathbf{y},i),0) d\mathbf{y}, \\
I_y^k &= \int_{\mathbf{y}\in N_{\mathbf{x}}} h_{\sigma^2(\mathbf{x},t)}(\mathbf{x},\mathbf{y}) \partial_y I(\mathbf{y}+\mathbf{w}(\mathbf{y},i),0) + \frac{\partial h_{\sigma^2(\mathbf{x},t)}(\mathbf{x},\mathbf{y})}{\partial s(\mathbf{x}+\mathbf{w}(\mathbf{x},i))} \partial_y s(\mathbf{x}+\mathbf{w}(\mathbf{x},i)) I(\mathbf{y}+\mathbf{w}(\mathbf{y},i),0) d\mathbf{y},
\end{aligned} \tag{19}$$

which can be solved by Gauss-Seidel or SOR iterations.

4.3 Depth Refinement

For the minimization of scene depth $s(\mathbf{x})$, we keep the distorted displacement $\mathbf{w}(\mathbf{y},t)$ fixed, then Eq. 13 becomes

$$\min_s J_2(s(\mathbf{x})) = \min_s E_d(s(\mathbf{x})) + \beta E_m(s(\mathbf{x})), \tag{20}$$

where $J_2(s(\mathbf{x}))$ is the depth refinement objective function; $E_d(s(\mathbf{x}))$ is the data-term defined in Eq. 14 with fixed $\mathbf{w}(\mathbf{y},t)$; $E_m(s(\mathbf{x}))$ is the depth regularization term defined in Eq. 16.

The minimization in Eq. 20 resembles the DFD optimization problem in the air [12,21], except that defocused blur kernel size functions (Eq. 8) and the $\psi(\xi^2)$ function are different. Thus, we modify the numerical solution in [21] for our problem.

4.4 Recover Surface Shape

In Sec. 4.1 to 4.3, we have proposed the optimization method to estimate the distortion between the reference image and refractive blur video and the underneath scene depth in the reference image. Based on Eq. 8 and Eq. 9, the normal vectors map of the wavy water surface in each frame can be reconstructed by the following linear operator:

$$\begin{cases}
f_x(\mathbf{x_a},t) = -\dfrac{n_x(\mathbf{x},t)}{n_z(\mathbf{x},t)} = -\dfrac{s(\mathbf{x}+\mathbf{w}(\mathbf{x},i)) + (n-1)H}{\kappa v_0 (n-1)(s(\mathbf{x}+\mathbf{w}(\mathbf{x},i)-H)} u(\mathbf{x},t) \\
f_y(\mathbf{x_a},t) = -\dfrac{n_y(\mathbf{x},t)}{n_z(\mathbf{x},t)} = -\dfrac{s(\mathbf{x}+\mathbf{w}(\mathbf{x},i)) + (n-1)H}{\kappa v_0 (n-1)(s(\mathbf{x}+\mathbf{w}(\mathbf{x},i)-H)} v(\mathbf{x},t),
\end{cases} \tag{21}$$

where $f_x(\mathbf{x_a},t)$ and $f_y(\mathbf{x_a},t)$ are the x-axis and y-axis gradients of wavy water surface at the point $\mathbf{x_a}$ corresponding to the pixel $\mathbf{x}$ in frames t. Surface integration from a gradient field by solving the Poisson equation has been well studied. In this paper, we apply a similar approach in [2,1] to recover the wavy water surface from its gradients field obtained from Eq. 21.

5 Experimental Results

5.1 Synthetic Data

We evaluate the performance of our approach on synthetic image sequences generated by ray-tracing method. Firstly the ground truth AIF image is regarded as the reference image taken through flat water surface. Its pixels are projected into the scene immersed in water by tracking light through the flat interface. Then under wavy interface, we track each light ray emitted from the scene pixels back to search for each pixels' new locations and refractive blur size on the sensor. Finally we render the synthetic image sequences based on Eq. 2 in Sec. 3.1.

We use the "cloth", "barn" and "baby" data from the Middlebury Stereo Datasets [16,27] and downsample them to resolution 148×174 pixels as the ground truth AIF image and depth. We assume that the distance from the camera to the water surface is 1m, the water depth is 0.5m, the depth range of object under the water is 1.1∼1.5m (we linearly map the ground truth depth to this range under the water), the refractive index of water is 1.33, the virtual camera is focused on the bottom of the water, the camera's focal length $f = 35mm$, the $f/\# = 4$ and the calibration parameter $\kappa = 3e4$. In these experiments, we implement our method with Matlab on a PC with an Intel 2.50G Hz Xeon Quad-Core E5420 CPU.

We first conduct experiments assuming the water surface to be a sinusoidal wave $z(x, y, t) = -0.5 + 0.001\cos(\pi t\sqrt{x^2 + y^2}/300)$ meters (we call this wave wave 1 in this paper, the coordinate system is the same with Sec. 3.2, as shown in Fig. 3(a)). The computation time is about 120 minutes (17 frames,"cloth" data), 80 minutes (15 frames,"barn" data) and 110 minutes (15 frames,"baby" data), respectively. For each scene, we pick one frame from synthetic video and its reconstruction results (depth and fluid surface shape) as shown in Fig. 4 (the first row in each subgraph), and the fluid surface reconstruction error maps are visualized as well (sixth column in Fig. 4). We can see the accuracy of the recovered scene depth and dynamic surfaces results with comparison to the ground truth data. For quantitative evaluation, we also calculate the root mean square error (RMSE) of the immersed scene's depth and water surface's shape sequence. The RMSE of the reconstructed scene depths are respectively 0.0555m("cloth"), 0.0358m("barn") and 0.0839m("baby"), and the average RMSE of the recovered water surfaces are 0.0362mm("cloth"), 0.0317mm("barn") and 0.0618mm ("baby").

To verify our approach's robustness towards different water fluctuations, we apply a different synthetic sinusoidal wave $z(x, y, t) = -0.5 + 0.001\cos(\pi x/60 + 9\pi t/32)$m to three scenes repeatedly (we call this wave wave 2 in this paper), and display the results in second row of each Fig. 4's subgraph. The computation time is about 120 minutes ("cloth"), 90 minutes ("barn") and 130 minutes ("baby") on 17 frames, respectively. The RMSE of the reconstructed scene depth are 0.0441m("cloth"), 0.0433m("barn") and 0.0767m("baby"), and the average RMSE of the reconstructed surfaces are 0.0648mm("cloth"), 0.0677mm("barn") and 0.1426mm("baby"). The results show similar accuracy to that in the first

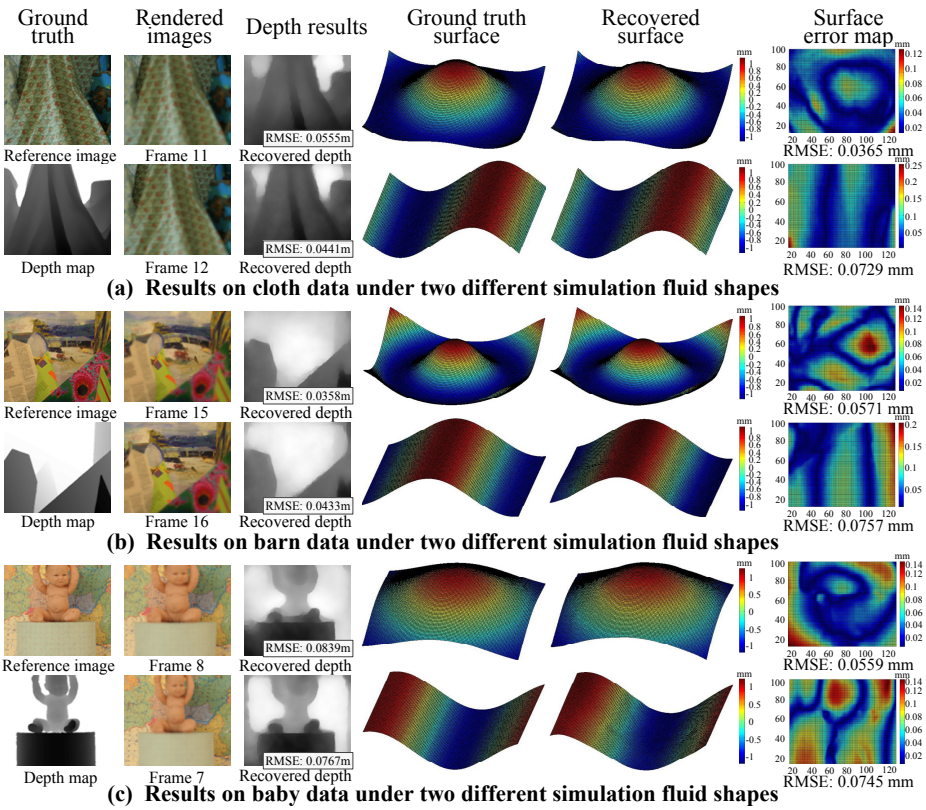
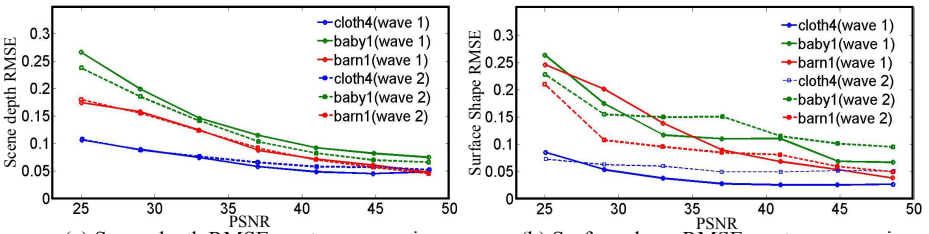


Fig. 4. Experimental results on three different synthetic data under two kinds of sinusoid waves. We synthesize the refractive blur images (second column) from reference image and its depth (first column) with known sinusoid waves (fourth column), then the depth and water surface are reconstructed (third and fifth column) with water surface error map (sixth column).

surface shape sequences. The complete recovered sequences can be found in the supplementary video. We also conduct similar experiments with the estimated reference reconstructed from the synthetic distorted AIF video by [26], of which results are shown in the supplementary material and video.

We also demonstrate the sensitivity of the proposed method by introducing different levels of additive white gaussian noises to the input video and the reference image. In implementation, Fig. 5 demonstrates the performance (RMSE of the reconstruction result) at varying noise levels. The curves show that the performance does not degenerate largely at increasing noise, especially on the rich-textured scene—'cloth', the RMSE on which is relatively lower than on the other scenes.



(a) Scene depth RMSE w.r.t. sensor noise (b) Surface shape RMSE w.r.t. sensor noise

Fig. 5. The sensitivity of the proposed method to sensor noise on three scenes under two different sinusoid waves. (a) The performance of estimated depth w.r.t. sensor noise. (b) The performance of estimated fluid surface shape w.r.t. sensor noise.

Box data Toy data Fish data

Fig. 6. The captured reference images of three real scenes

5.2 Real-captured Results

In order to reconstruct the real fluid surface and depth of the underneath scene, we set up a system as shown in Fig. 3(b). The camera (Canon EOS 5DII) is placed orthogonally to the flat water surface. The scene is uniformly lit from the side face of tank to avoid specular reflections. The distance from the camera to the bottom of the tank is about 72 cm and the depth of water is about 40 cm.

To test our approach's performance under a variety of conditions, we conduct a series of experiments on three immersed scenes with different texture richness of which reference images are shown in Fig. 6 and two different kinds of water fluctuation amplitudes. We firstly take an AIF image through the undisturbed water surface with a small aperture directly to avoid the inaccuracy of the reconstructed reference image by using [30,26], and regard it as the reference image in our approach. Then we capture a blurry video through wavy water with a large aperture, and apply our approach to above two inputs.

Firstly we test our approach under a slightly rippled water surface generated by dripping a drop of water into the tank. To eliminate influences on the video capturing, we drop the water at the corner and generate quarter-annular waves. The results in the first row of each Fig. 7's subgraph show that we achieve promising reconstruction in scenes with abundant texture (Fig. 7(a), with $f = 35mm$ and $f/\# = 4$), common texture scene (Fig. 7(b), with $f = 50mm$ and $f/\# = 4$) and textureless scene (Fig. 7(c), with $f = 50mm$ and $f/\# = 4$).

We also test our method under larger fluctuating water surface by blowing air onto it using a hair dryer. We repeat the same process to capture the reference image and refractive blurry video and reconstruct both the water surface shape

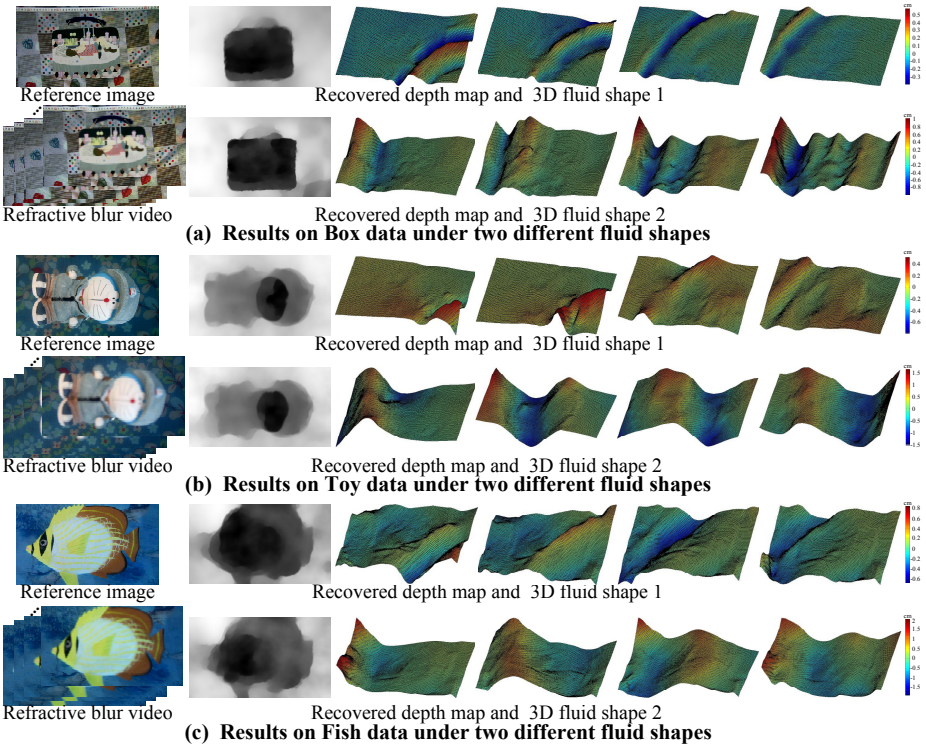


Fig. 7. Results on real-captured data. We reconstruct the depth (second column) and water surface (third to sixth columns) from a captured refractive blur video and a reference image (first column).

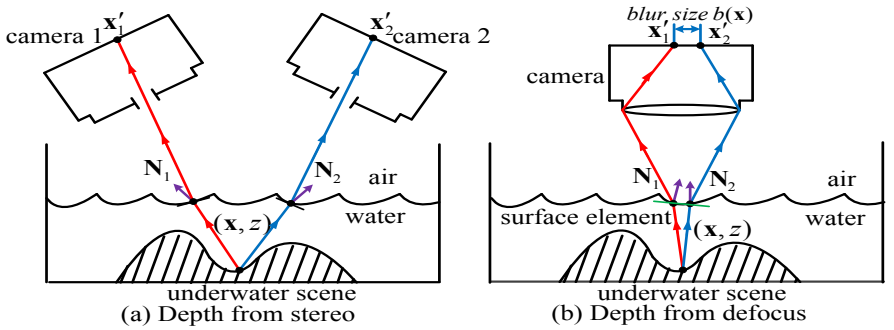


Fig. 8. Comparison of the stereo (a) and DFD approaches (b) applied for scenes under non-planar fluid surface. In stereo, rays leaving the same scene point crosses the fluid surface at different locations, which may have different normals. Compared to stereo, in DFD, two normal vectors of the refractive plane can be regarded as the same due to small baseline (aperture size).

and underneath scene depth. The second row in each Fig. 7's subgraph shows the reconstruction results on the three scenes and demonstrates the effectiveness of the proposed approach in such cases.

6 Discussion and Conclusions

This paper has presented the first method that exploits defocus to simultaneously reconstruct dynamic fluid surface and depth of the immersed scene using a single camera. We build a refractive blur geometry model and develop an iterative inference algorithm for depth and surface shape recovery. The performance and robustness of our approach are experimentally validated on both synthetic and real data.

Comparison with Stereo. Compared with other 3D reconstruction methods such as stereo for imaging through wavy fluid surface, there are some advantages of our method:

- As shown in Fig.8(b) and mentioned in Sec. 3.3, the normal vectors in the surface element can be assumed to be the same in our model owing to the small baseline (aperture size). Thus compared with Fig.8(a), there is less number of the unknowns in the geometry, which enhances the performance of our reconstruction algorithms.
- Similar to DFD method in the air, our approach uses only single camera, avoids multi-view registration and is robust to occlusion, as analyzed in [29]. Besides, the geometry registration in DFD method can be eliminated since we only change the aperture size during capturing and there is no scaling between images. The proposed model incorporates multiple controllable camera parameters (e.g., aperture diameter, focusing depth) and thus is flexible for developing high performance algorithms.

Although our approach could achieve promising performance on various scenes under different types of fluid surfaces, using multiple cameras or changing camera settings to further enhance the reconstruction accuracy are two interesting avenues of future research. Another interesting future direction is to handle the absorption and scattering of light in media such as fog, smoke and murky water[25,15,28].

Acknowledgments. This work was supported by the Project of NSFC (No.61327902, No.61035002 and No.61120106003), and Mohit was supported by NSF grant IIS 09-64429.

References

1. Agrawal, A., Chellappa, R., Raskar, R.: An algebraic approach to surface reconstruction from gradient fields. In: ICCV, vol. 1, pp. 174–181. IEEE (2005)
2. Agrawal, A., Raskar, R., Chellappa, R.: What is the range of surface reconstructions from a gradient field? In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 578–591. Springer, Heidelberg (2006)
3. Alterman, M., Schechner, Y., Perona, P., Shamir, J.: Detecting motion through dynamic refraction. PAMI 35(1), 245–251 (2013)
4. Alterman, M., Schechner, Y.Y., Swirski, Y.: Triangulation in random refractive distortions. In: ICCP, pp. 1–10. IEEE (2013)
5. Ben-Ezra, M., Nayar, S.K.: What does motion reveal about transparency? In: ICCV, pp. 1025–1032. IEEE (2003)
6. Brox, T., Bregler, C., Malik, J.: Large displacement optical flow. In: CVPR, pp. 41–48. IEEE (2009)
7. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3024, pp. 25–36. Springer, Heidelberg (2004)
8. Chang, Y.-J., Chen, T.: Multi-view 3d reconstruction for scenes under the refractive plane with known vertical direction. In: ICCV, pp. 351–358. IEEE (2011)
9. Ding, Y., Li, F., Ji, Y., Yu, J.: Dynamic fluid surface acquisition using a camera array. In: ICCV, pp. 2478–2485. IEEE (2011)
10. Donate, A., Ribeiro, E.: Improved reconstruction of images distorted by water waves. In: Advances in Computer Graphics and Computer Vision, pp. 264–277. Springer, Heidelberg (2007)
11. Efros, A., Isler, V., Shi, J., Visontai, M.: Seeing through water. In: NIPS, vol. 17, pp. 393–400 (2005)
12. Favaro, P.: Recovering thin structures via nonlocal-means regularization with application to depth from defocus. In: CVPR, pp. 1133–1140. IEEE (2010)
13. Favaro, P., Soatto, S.: 3D shape reconstruction and image restoration: exploiting defocus and motion blur. Springer Verlag (2006)
14. Ferreira, R., Costeira, J.P., Santos, J.A.: Stereo reconstruction of a submerged scene. In: Marques, J.S., Pérez de la Blanca, N., Pina, P. (eds.) IbPRIA 2005. LNCS, vol. 3522, pp. 102–109. Springer, Heidelberg (2005)
15. Gupta, M., Narasimhan, S.G., Schechner, Y.Y.: On controlling light transport in poor visibility environments. In: CVPR, pp. 1–8. IEEE (2008)
16. Hirschmuller, H., Scharstein, D.: Evaluation of cost functions for stereo matching. In: CVPR, pp. 1–8. IEEE (2007)
17. Huynh, C.P., Robles-Kelly, A., Hancock, E.: Shape and refractive index recovery from single-view polarisation images. In: CVPR, pp. 1229–1236. IEEE (2010)
18. Ihrke, I., Goidluecke, B., Magnor, M.: Reconstructing the geometry of flowing water. In: ICCV, vol. 2, pp. 1055–1060. IEEE (2005)
19. Jähne, B., Klinke, J., Waas, S.: Imaging of short ocean wind waves: a critical theoretical review. JOSA A 11(8), 2197–2209 (1994)
20. Kidger, M.J.: Fundamental optical design, vol. 92. SPIE Press Bellingham, Washington, DC (2002)
21. Lin, X., Suo, J., Cao, X., Dai, Q.: Iterative feedback estimation of depth and radiance from defocused images. In: Lee, K.M., Matsushita, Y., Rehg, J.M., Hu, Z. (eds.) ACCV 2012, Part IV. LNCS, vol. 7727, pp. 95–109. Springer, Heidelberg (2013)

22. Morris, N.J., Kutulakos, K.N.: Dynamic refraction stereo. In: ICCV, vol. 2, pp. 1573–1580. IEEE (2005)
23. Morris, N.J., Kutulakos, K.N.: Reconstructing the surface of inhomogeneous transparent scenes by scatter-trace photography. In: ICCV, pp. 1–8. IEEE (2007)
24. Murase, H.: Surface shape reconstruction of an undulating transparent object. In: ICCV, pp. 313–317. IEEE (1990)
25. Narasimhan, S.G., Nayar, S.K., Sun, B., Koppal, S.J.: Structured light in scattering media. In: ICCV, vol. 1, pp. 420–427. IEEE ((2005)
26. Oreifej, O., Shu, G., Pace, T., Shah, M.: A two-stage reconstruction approach for seeing through water. In: CVPR, pp. 1153–1160. IEEE (2011)
27. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. IJCV 47(1-3), 7–42 (2002)
28. Schechner, Y.Y., Karpel, N.: Recovery of underwater visibility and structure by polarization analysis. IEEE Journal of Oceanic Engineering 30(3), 570–587 (2005)
29. Schechner, Y.Y., Kiryati, N.: Depth from defocus vs. stereo: How different really are they? IJCV 39(2), 141–162 (2000)
30. Tian, Y., Narasimhan, S.G.: Seeing through water: Image restoration using model-based tracking. In: ICCV, pp. 2303–2310. IEEE (2009)
31. Tian, Y., Narasimhan, S.G.: A globally optimal data-driven approach for image distortion estimation. In: CVPR, pp. 1277–1284. IEEE (2010)
32. Treibitz, T., Schechner, Y., Kunz, C., Singh, H.: Flat refractive geometry. PAMI 34(1), 51–65 (2012)
33. Wen, Z., Lambert, A., Fraser, D., Li, H.: Bispectral analysis and recovery of images distorted by a moving water surface. Applied Optics 49(33), 6376–6384 (2010)
34. Wetzstein, G., Roodnick, D., Heidrich, W., Raskar, R.: Refractive shape from light field distortion. In: ICCV, pp. 1180–1186. IEEE (2011)
35. Yau, T., Gong, M., Yang, Y.-H.: Underwater camera calibration using wavelength triangulation. In: CVPR, pp. 2499–2506. IEEE (2013)
36. Ye, J., Ji, Y., Li, F., Yu, J.: Angular domain reconstruction of dynamic 3d fluid surfaces. In: CVPR, pp. 310–317. IEEE (2012)