

Triangulation in Random Refractive Distortions

Marina Alterman, Yoav. Y. Schechner and Yohay Swirski
Dept. of Electrical Engineering,
Technion - Israel Institute of Technology, Haifa, Israel

amarina@tx.technion.ac.il, yoav@ee.technion.ac.il, yohays@tx.technion.ac.il

Abstract

Random refraction occurs in turbulence and through a wavy water-air interface. It creates distortion that changes in space, time and with viewpoint. Localizing objects in three dimensions (3D) despite this random distortion is important to some predators and also to submariners avoiding the salient use of periscopes. We take a multiview approach to this task. Refracted distortion statistics induce a probabilistic relation between any pixel location and a line of sight in space. Measurements of an object's random projection from multiple views and times lead to a likelihood function of the object's 3D location. The likelihood leads to estimates of the 3D location and its uncertainty. Furthermore, multiview images acquired simultaneously in a wide stereo baseline have uncorrelated distortions. This helps reduce the acquisition time needed for localization. The method is demonstrated in stereoscopic video sequences, both in a lab and a swimming pool.

1. Introduction

Random refraction in visual sensing is caused by thermal turbulence in the atmosphere and deep-water volcanic vents. This creates a distortion field that is random in space and time. More strongly, this effect occurs when looking through a wavy water-air interface (WAI), either into water from airborne positions, or out to the air from submerged viewpoints. The latter case is most severe, since small changes in a WAI slope lead to large angular changes of an airborne line of sight (LOS), due to Snell's law. Example images are shown in Fig. 1.

There are biological motivations and engineering applications to study vision in such scenarios, particularly in multiview settings. As illustrated in Fig. 2a, birds [25] fly in search of submerged fish to hunt. Some animals hunt the other way. For example, the archer fish [3] launches water jets in a parabolic ballistic trajectory to shoot down flies (Fig. 2b) that sit on foliage. In all cases, predators need to assess well the 3D location of the prey prior to charging.

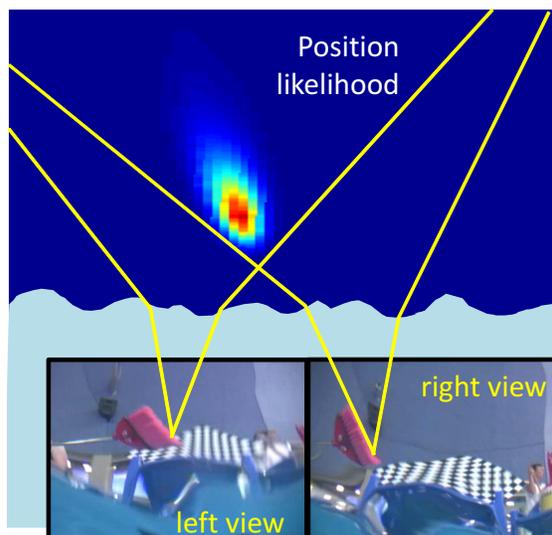


Figure 1. Stereo pair sample images taken by looking upward through a wavy water surface. By triangulation using our method we obtain a likelihood function in 3D. The estimated location is the position of the maximum likelihood (ML).

This can be done by triangulation as the predator changes its position during its path, thus obtaining multi-views. Upward vision through a wavy WAI can serve as a *virtual periscope* (Fig. 2c). This can help submariners assess activities above water, without using physical periscopes which flag their presence.

In open air, three dimensional (3D) scene reconstruction using multiple views is well studied [21, 37, 44, 54]. Distortions through a *flat* WAI have also been thoroughly studied [7, 8, 16, 17, 19, 28, 39, 53]. Measuring from air the 3D structure of submerged scenes has been proposed based on stereo [16] or motion [7, 39]. Still, a triangulation challenge remains when a flat-WAI is perturbed by random unknown WAI waves. This paper addresses this challenge.

There are video-based methods to “flatten” monocular images taken through a wavy WAI. Lucky imaging [14, 15, 46, 55] require a large number of frames, out which

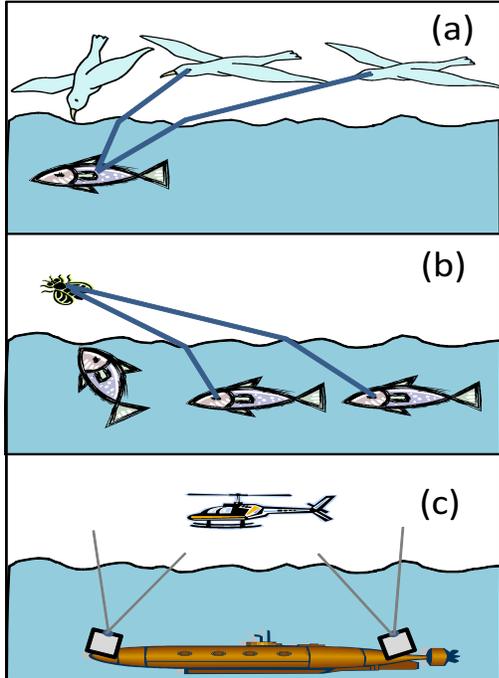


Figure 2. Relevant scenarios. (a) Birds fly by to hunt for a submerged fish. (b) The archer fish shoots water jets to shoot down airborne flies. (c) A submarine avoiding physical periscope by upward vision through a wavy WAI.

a best-representative frame is selected per patch. Less temporal frames are needed if a spatiotemporal distortion model [34, 35, 51] is fitted to the image data, or given an undistorted template image [52]. A virtual periscope is theoretically proposed [30] based on a wave model constrained by self occlusions. Possibly, rectification can be based on WAI estimates derived from dedicated optical measurements [13, 22, 43], some of which need structured light [2, 10]. Ref. [1] seeks motion detection through a wavy WAI. There, a motion classifier is used, based on statistical properties of water waves, without needing image rectification.

Monocular methods as mentioned above do not explicitly triangulate objects and measure their range. However, they can be useful: if rectified images are estimated in a first step, they can be triangulated in a second step.¹ Thus, some elements of these prior studies can help improve the method we propose. Nevertheless, there are benefits in dealing *directly* with raw multiview data. First, random errors in this difficult rectification task generally make deterministic triangulation very difficult, since erroneous LOSs may not intersect in 3D. Second, a “flattened” monocular image may harbor a global bias: a very smooth but slanted

¹In [50], a single statistic, the mean range through atmospheric turbulence, was derived based on mean pixel projection per view.

WAI gives correct image content but induces a spatial offset that biases triangulation. Third, most importantly, *widening the viewpoints* support *shortens the acquisition time* required for high quality object localization. The reason is that images taken *in parallel* from distant locations have uncorrelated distortions. Thus, instead of using more time to sequentially acquire new uncorrelated frames, we use the spatial dimension of viewpoint location to obtain the data.²

We note that there are multi-camera methods dedicated to recovering a wavy WAI [12, 33, 41, 42, 56]. These methods, however, often rely on a known calibration target being the observed object. They are thus not intended to triangulate objects behind a WAI.

We take a stochastic approach to triangulation under random refractive distortions. The statistics of WAI waves induce by refraction a probabilistic relation between any pixel location and an airborne LOS. Measurements of an object’s random projection from multiple views and times lead to a likelihood function of the object’s 3D location.³ Maximum likelihood (ML) estimates the 3D location, while the effective support of the likelihood function informs of the location uncertainty.

2. Theoretical Background

Refraction has been used and analyzed in the context of computational photography [20, 31, 33, 40, 47, 48, 49]. Theory about visual refraction through a WAI is described in Ref. [1]. This section briefly follows notations and relevant derivations from Ref. [1].

2.1. Snell Law in Various Forms

Consider Fig. 3. Let us back-project a LOS from a submerged camera towards an airborne object point. The LOS in water is directed along unit vector $\hat{\mathbf{v}}_w$. This vector forms an angle θ_w with the WAI normal $\hat{\mathbf{N}}$, as illustrated in Fig. 3. At the WAI, the back-projected LOS refracts and proceeds in air along unit vector $\hat{\mathbf{v}}_a$. This vector forms an angle θ_a with $\hat{\mathbf{N}}$. The scalar Snell law of refraction is

$$\sin \theta_a = n \sin \theta_w, \quad (1)$$

i.e.,

$$\cos \theta_a = \sqrt{1 - n^2 + n^2 \cos^2 \theta_w}. \quad (2)$$

Vector forms of Snell’s law [26] are

$$\hat{\mathbf{v}}_a \times \hat{\mathbf{N}} = n \hat{\mathbf{v}}_w \times \hat{\mathbf{N}}. \quad (3)$$

$$\hat{\mathbf{v}}_a = n \hat{\mathbf{v}}_w + \hat{\mathbf{N}}(\cos \theta_a - n \cos \theta_w), \quad (4)$$

²In atmospheric turbulence, a wide camera aperture reduces the chance for deconvolution-favorable imaging. Hence Ref. [29] suggests splitting the aperture into subapertures, and making parallel image acquisitions.

³Dellaert *et al.* [11] formulates an iterative algorithm to obtain the maximum likelihood 3D structure and camera motion given image measurements.

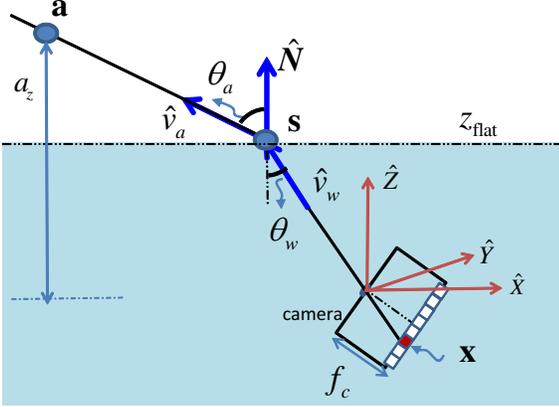


Figure 3. Viewing geometry through a flat water surface.

where

$$\cos \theta_w = \hat{\mathbf{v}}_w \cdot \hat{\mathbf{N}}. \quad (5)$$

Here \times is the cross product and $n \approx 1.33$ is the optical refractive index of water. From Eq. (3), $\hat{\mathbf{v}}_w$, $\hat{\mathbf{v}}_a$ and $\hat{\mathbf{N}}$ are co-planar (in the plane of incidence).

2.2. Derivation of Back-Projection

A submerged camera has an internal 3D coordinate system. Its origin is at the center of projection. The axes include the optical axis and the lateral pixel coordinates of the image plane. A pixel length is l^{pixel} . The optical axis intersects the image plane at pixel \mathbf{c} . In the camera coordinate system, pixel \mathbf{x} is at physical location

$$\mathbf{x}^{\text{cam}} = \begin{bmatrix} \mathbf{x} - \mathbf{c} \\ f_c \end{bmatrix} l^{\text{pixel}}, \quad (6)$$

where f_c is the focal length of the camera in units of pixels.

The origin of the global (lab) 3D coordinate system is set to be also at the center of projection. The global coordinates are composed of the zenith axis $\hat{\mathbf{Z}}$ and two horizontal coordinates, $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$. In the global coordinate system, the 3D location of pixel \mathbf{x} is

$$\mathbf{x}_{\text{lab}} = \mathbf{R}^T(\mathbf{x}^{\text{cam}}), \quad (7)$$

where \mathbf{R} is the camera rotation matrix and T denotes transposition. Thus, in the lab coordinate system, the submerged LOS is along unit vector

$$\hat{\mathbf{v}}_w = \mathbf{x}_{\text{lab}} / \|\mathbf{x}_{\text{lab}}\|. \quad (8)$$

We can substitute Eqs. (5) and (2) into Eq. (4), and set $\hat{\mathbf{N}} = \hat{\mathbf{Z}}$ to express imaging through a flat WAI. This yields the airborne LOS direction

$$\hat{\mathbf{v}}_a = n\hat{\mathbf{v}}_w + \hat{\mathbf{Z}} \left[\sqrt{1 - n^2 + n^2(\hat{\mathbf{v}}_w \cdot \hat{\mathbf{Z}})^2} - n\hat{\mathbf{v}}_w \cdot \hat{\mathbf{Z}} \right] \quad (9)$$

The vertical components of $\hat{\mathbf{v}}_w$ and $\hat{\mathbf{v}}_a$ are respectively denoted by $v_{w,z}$ and $v_{a,z}$.

An airborne object is at $\mathbf{a} = (a_x, a_y, a_z)^T$ in the global coordinate system (Fig. 3). Back-propagating from the camera's center of projection, a refracted LOS includes a submerged LOS segment and an airborne one. Thus,

$$\mathbf{a} = \hat{\mathbf{v}}_w l_w + \hat{\mathbf{v}}_a l_a, \quad (10)$$

where l_w and l_a are line-length parameters. Suppose the camera is at depth z_{flat} below a flat WAI. Then,

$$l_w = z_{\text{flat}}/v_{w,z}, \quad l_a = (a_z - z_{\text{flat}})/v_{a,z}. \quad (11)$$

The parameters z_{flat} and \mathbf{R} can be measured by the viewing system. A depth gauge in the system can measure z_{flat} . The orientation of the cameras can be obtained by an integrated accelerometer as in a smartphone, which has been used in computer vision to address refractive distortions [7]. In nature, fish can control their balance using otoliths [27] or a swim-bladder [45]. Even without an accelerometer, the inclination angle can be computed by the location of Snell's window in the field of view [1].

3. Modeling Position Statistics

3.1. Random Projection Distribution: Single View

Consider Fig. 3. When the WAI is flat, a projects to pixel \mathbf{x}_{flat} . When the WAI is wavy, a projects to pixel $\mathbf{x}(t)$ at time t , where

$$\mathbf{x}(t) = \mathbf{x}_{\text{flat}} + \mathbf{d}(\mathbf{x}_{\text{flat}}, t). \quad (12)$$

Here \mathbf{d} is the displacement in the image of the object, caused by the random WAI waves. In other words, the spatio-temporal field $\mathbf{d}(\mathbf{x}_{\text{flat}}, t)$ is a *random distortion* created by random refractive changes. Following Cox and Munk [9], the WAI normal $\hat{\mathbf{N}}$ is random in spacetime and has a Gaussian distribution. The variance of $\hat{\mathbf{N}}$ depends on meteorological parameters. For a given \mathbf{x}_{flat} , the random vector \mathbf{d} has approximately a normal [1] distribution: $\mathbf{d} \sim \mathcal{N}(0, \Sigma_{\mathbf{x}})$. The 2×2 covariance matrix $\Sigma_{\mathbf{x}}$ depends on the WAI roughness, the camera parameters (\mathbf{R} , l^{pixel} , f_c) and somewhat on \mathbf{x}_{flat} , as described in [1]. Thus, the probability density function (PDF) of imaging a at \mathbf{x} is

$$p(\mathbf{x}|\mathbf{x}_{\text{flat}}) \approx G \exp \left[-\frac{1}{2}(\mathbf{x} - \mathbf{x}_{\text{flat}})^T \Sigma_{\mathbf{x}}^{-1}(\mathbf{x} - \mathbf{x}_{\text{flat}}) \right], \quad (13)$$

where G is a normalization factor.

In our case, \mathbf{x}_{flat} is not given. Thus, according to Bayes' rule, the probability of \mathbf{x}_{flat} given \mathbf{x} is,

$$p(\mathbf{x}_{\text{flat}}|\mathbf{x}) \propto p(\mathbf{x}|\mathbf{x}_{\text{flat}})p(\mathbf{x}_{\text{flat}}). \quad (14)$$

The PDF $p(\mathbf{x}_{\text{flat}})$ is a prior on where the object might preferably be projected to, when the WAI is flat. Usually,

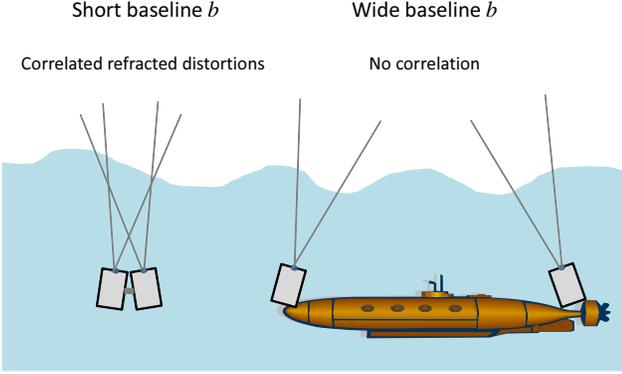


Figure 4. [Left] Baseline that is significantly shorter than the typical correlation length of the WAI-slope, results in highly correlated distortions across views. [Right] A wide baseline results in uncorrelated distortions between the two views.

there is no such preference. Hence $p(\mathbf{x}_{\text{flat}})$ is a constant, and

$$p(\mathbf{x}_{\text{flat}}|\mathbf{x}) \sim p(\mathbf{x}|\mathbf{x}_{\text{flat}}). \quad (15)$$

3.2. Uncorrelated Multi-Views

We now focus on random refraction in multiview geometry. Consider a submerged stereo system having a baseline b . Variables associated with the left or right camera are denoted with L or R, respectively. Each camera views the object through a different WAI portion: one is around horizontal location $\mathbf{s}^L = (s_x^L, s_y^L)$, while the other is around $\mathbf{s}^R = (s_x^R, s_y^R)$. The projected image point on the left is distorted (displaced) by \mathbf{d}^L , while the corresponding point on the right image is displaced by \mathbf{d}^R .

Is the random dynamic distortion $\mathbf{d}^L(t)$ similar or very different than $\mathbf{d}^R(t)$, $\forall t$? How mutually statistically dependent are \mathbf{d}^L and \mathbf{d}^R ? The distortion mainly depends on the WAI slope [1], i.e., on $\dot{\mathbf{N}}(\mathbf{s}, t)$. Hence, statistical dependency between \mathbf{d}^L and \mathbf{d}^R is dictated by the dependency between $\dot{\mathbf{N}}(\mathbf{s}^L)$ and $\dot{\mathbf{N}}(\mathbf{s}^R)$. According to Ref. [9], the WAI slope statistics is approximately Gaussian. Hence, statistical dependency of slopes is equivalent to correlation. The correlation between slopes⁴ in \mathbf{s}^L and \mathbf{s}^R dictates the correlation between \mathbf{d}^L and \mathbf{d}^R .

Correlation between WAI slopes at \mathbf{s}^L and \mathbf{s}^R depends on the lateral distance $\|\mathbf{s}^L - \mathbf{s}^R\| \approx b$. If the baseline b is significantly shorter than the typical WAI-slope *correlation length*, then distortions in the two views are mutually highly correlated, as illustrated in Fig. 4.

We prefer to work in another domain, where distortions in the multiple views are mutually *uncorrelated*, largely. Low statistical dependency (low mutual information) be-

⁴A model to characterize WAI slope and curvature variances and the spatial correlation of slopes, was proposed in [32].

tween multiview measurements means that any new view adds *more information* about the object location. This is achieved if b is significantly larger than the typical WAI-slope correlation length (See Fig. 4). In this context, thus, the wider the baseline, the better. We do not know the WAI-slope correlation length a-priori. However, in our experiments, we verified empirically the low correlation between \mathbf{d}^L and \mathbf{d}^R .

3.3. Airborne Position Likelihood

Single View

Under a flat WAI, the object in \mathbf{a} projects to pixel $\mathbf{x}_{\text{flat}}^L$ in camera L. Through a flat WAI, there is one-to-one correspondence between $\mathbf{x}_{\text{flat}}^L$ and a specific LOS, denoted $\text{LOS}(\mathbf{x}_{\text{flat}}^L)$, by back-projection. Hence, any probability density associated with $\mathbf{x}_{\text{flat}}^L$ is also associated with $\text{LOS}(\mathbf{x}_{\text{flat}}^L)$.

Since the WAI is wavy, \mathbf{a} projects to pixel $\mathbf{x}^L(t)$, at time t , while $\mathbf{x}_{\text{flat}}^L$ is unknown. Eq. (15) sets a probability density $p[\mathbf{x}_{\text{flat}}^L|\mathbf{x}^L(t)]$ to each pixel $\mathbf{x}_{\text{flat}}^L$ in the L image plane. Thus, Eqs. (13,15) set a probability density

$$p[\text{LOS}(\mathbf{x}_{\text{flat}}^L)|\mathbf{x}^L(t)] \sim p[\mathbf{x}^L(t)|\mathbf{x}_{\text{flat}}^L]. \quad (16)$$

So $\forall \mathbf{x}_{\text{flat}}^L$, Eq. (16) back-projects an image-domain PDF to a PDF of all LOSs that can backproject from camera L through a flat WAI.

An LOS is an infinite set of 3D points \mathbf{X} that project to the same image point. A priori, each of these 3D points is equally likely to be the sought object \mathbf{a} . Hence, with any point $\mathbf{X} \in \text{LOS}(\mathbf{x}_{\text{flat}}^L)$, we associate a *likelihood* equivalent to the probability density defined in Eq. (16):

$$l_t^L(\mathbf{X}) \equiv p[\text{LOS}(\mathbf{x}_{\text{flat}}^L)|\mathbf{x}^L(t)] \quad | \quad \mathbf{X} \in \text{LOS}(\mathbf{x}_{\text{flat}}^L). \quad (17)$$

Based on Eqs. (16,17)

$$l_t^L(\mathbf{X}) \sim p[\mathbf{x}^L(t)|\mathbf{x}_{\text{flat}}^L] \quad | \quad \mathbf{X} \in \text{LOS}(\mathbf{x}_{\text{flat}}^L). \quad (18)$$

We work with discrete, regular spatial grids: image locations $\mathbf{x}^L(t)$, $\mathbf{x}_{\text{flat}}^L$ are integer pixel locations, and so are the 3D candidate object locations (voxels). Thus in practice, we calculate *samples* of the image-domain PDF, which correspond to sample backprojected LOSs. Hence, Eq. (18) derives the likelihood in samples (a point cloud) in 3D. Between these cloud points, we interpolate $l_t^L(\mathbf{X})$, $\forall \mathbf{X}$.

Multi-View

We have defined above the variables associated with the L camera. Let us generalize the formulation to multiple views, particularly stereo. Under a flat WAI, the object in \mathbf{a} projects to pixel $\mathbf{x}_{\text{flat}}^R$ in camera R. Since the WAI is wavy, \mathbf{a} projects to pixel $\mathbf{x}^R(t)$, at time t , while $\mathbf{x}_{\text{flat}}^R$ is unknown.

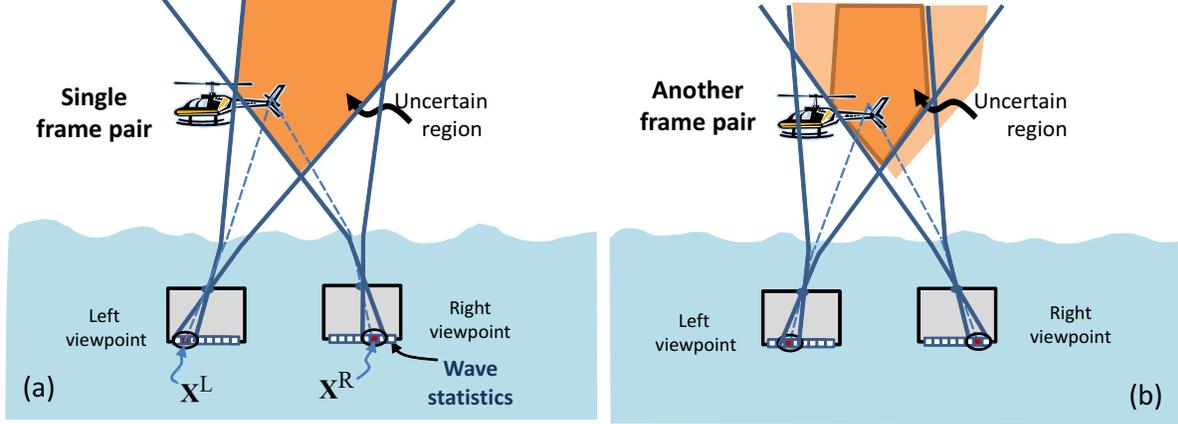


Figure 5. Airborne position likelihood. (a) By projecting the uncertainties around \mathbf{x}^L and \mathbf{x}^R , the effective 3D spatial support (orange region) of the overlap represents the 3D domain in which the airborne object point is likely to reside. (b) Additional temporal frames narrow this domain.

Similarly to the process involving Eq. (18), we derive an interpolated likelihood $l_t^R(\mathbf{X})$, $\forall \mathbf{X}$, based on

$$l_t^R(\mathbf{X}) \sim p[\mathbf{x}^R(t)|\mathbf{x}_{\text{flat}}^R] \mid \mathbf{X} \in \text{LOS}(\mathbf{x}_{\text{flat}}^R). \quad (19)$$

In this work we assume that between the views, correspondence of image points is established. In other words, we know that the measurement pixel set $\{\mathbf{x}^L(t), \mathbf{x}^R(t)\}_t$ corresponds to the same 3D object point, but we do not know where the object is.

Recall from Sec. 3.2 that distortions between views can be approximated as statistically independent. Therefore, likelihoods stemming from different frames and viewpoints multiply each other. Overall, for N_{frames} frames, the likelihood is

$$L(\mathbf{X}) = \prod_{t=1}^{N_{\text{frames}}} l_t^L(\mathbf{X}) l_t^R(\mathbf{X}), \quad (20)$$

as illustrated in Fig. 5.

The effective 3D spatial support (orange region in Fig. 5) of $L(\mathbf{X})$ represents the 3D domain in which the airborne object point is likely to reside. The more viewpoints and temporal frames, the narrower this domain becomes. Temporal frames can be traded-off for more viewpoints, in the quest to narrow the support of Eq. (20) by further likelihood factors. Hence, increasing the number of viewpoints can *shorten* the acquisition time needed for certain localization accuracy of the object. ML yields an estimate of an optimal airborne object location.

$$\hat{\mathbf{a}} = \underset{\mathbf{X}}{\text{argmax}} L(\mathbf{X}). \quad (21)$$

4. Experiments

4.1. Correlation

We conducted experiments in a lab’s water tank and in a swimming pool. We used a pair of Canon HV-30 camcorders, each in an underwater housing. Their baseline was $b = 27.5$ cm. Fig. 6 shows the lab setup. The video sequences of the two cameras were synchronized in post processing, by detecting light strobes that we had recorded. The rig was calibrated underwater (in a swimming pool) using a checkerboard target. Each of our camera housings has flat windows, which generally might invalidate the common single viewpoint model [8, 23, 53]. However, we used the Matlab calibration toolbox [4] both to calibrate and to verify measurements at other underwater distances. The results had negligible errors. Possibly this is due to the small angles we work in. To compute the effective range to objects, we use the pixel length l^{pixel} supplied by the camera manufacturer.

We used Large Displacement Optical Flow (LDOF) [5, 6] for tracking interest points⁵. We found this tracker to be rather robust to the harsh and fast distortions exhibited. First, we wanted to verify that distortions have just a small correlation between viewpoints. We tracked 40 points in the left and right frames. Then, we calculated the correlation coefficients between corresponding pixel locations. Distortion correlation was calculated in the components along the baseline (x axis) and perpendicular to the baseline. The respective average correlation coefficients are 0.12 and 0.16. This result points to low correlation between views. Then, we examined the temporal autocorrelation of the trajectory of each point. Autocorrelation decays in time. The effec-

⁵Recently, also a Locally Orderless Tracking (LOT) [36] demonstrated successful tracking under refractive distortions.

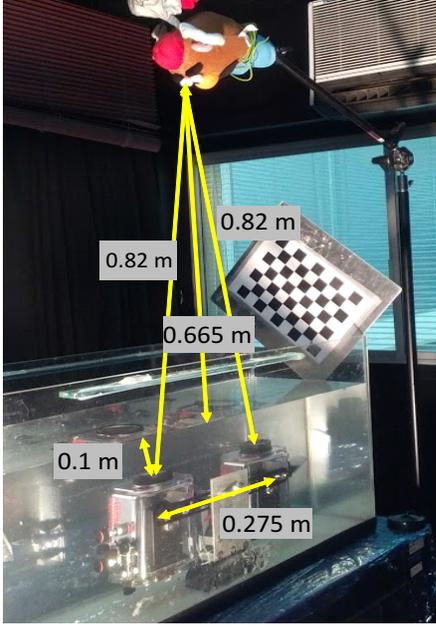


Figure 6. Lab experiment setup. Camcorders in a stereo rig are submerged in a water tank. Objects are hung in the air above, at known locations that serve as ground truth.

tive correlation time in our experiments was ≈ 9 frames, equivalent to $\approx 9/30 = 0.3$ second. Between consecutive frames, the correlation coefficient is 0.6. The conclusion is that correlation between viewpoints is low: significantly lower than temporal correlation in video. This strongly supports the use of multiview imaging, irrespective of triangulation, even if we only want to rectify distant object locations. Multiple views offer less correlated data (desirable), significantly faster than sequential monocular imaging.

We estimated $\Sigma_{\mathbf{x}}$, by fitting a Gaussian to trajectories of some tracked points, over several frames. The matrix is approximately diagonal when the camera's x axis is horizontal [1] in the lab coordinate⁶ system i.e. $\mathbf{x} \cdot (\hat{X} \times \hat{Y}) = 0$. In our case, this occurs when the baseline is horizontal. The standard deviations (STDs) in the respective image axes are σ_x and σ_y .

We opted not to rely solely on the Gaussian model for the distortion PDF $p(\mathbf{d})$. The reason is that sudden water movements and tracker errors can yield outlier location \mathbf{x} for a given \mathbf{x}_{flat} . To make the method robust to outliers, we incorporated a long tail into the modeled Gaussian, as illustrated in Fig. 7. This was achieved by setting the distortion PDF to be

$$p(\mathbf{d}) = \alpha \mathcal{N}(0, \Sigma_{\mathbf{x}}) + (1 - \alpha) \mathcal{N}(0, \mathbf{P}) \quad (22)$$

where \mathbf{P} is a diagonal matrix, expressing a Gaussian

⁶The camera's y axis can have an elevation angle in the lab coordinate system.

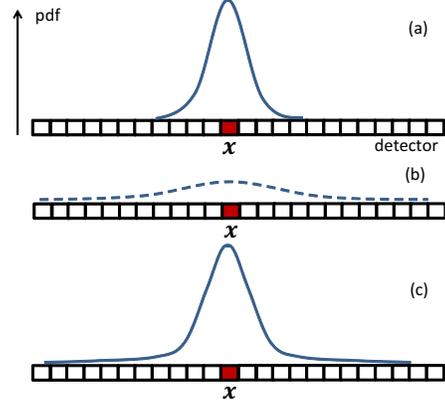


Figure 7. Pixel position statistical model. [a] Normal distribution of the distorted pixel location, around its undistorted location. [b] A long tail distribution, created by a Gaussian having a very wide support. [c] Weighted average of the distributions in [a] and [b].

whose axial STDs are respectively $7\sigma_x$ and $7\sigma_y$. We used $\alpha = 0.02$.

4.2. Laboratory Experiment

In the lab, the camera pair was placed in a water tank at a depth of $z_{\text{flat}} = 10\text{cm}$, looking upwards through the WAI (Fig. 6). The scene includes an object $\approx 82\text{cm}$ from the cameras. Its ground truth location in the lab coordinate system is $\mathbf{a} = [14, 8, 81]\text{cm}$. Sample pair frames are shown in Fig. 8a,b. Using the presented approach, the 3D object position was estimated to be $\hat{\mathbf{a}} = [16 \pm 2, 8 \pm 2, 80 \pm 18]\text{cm}$, as shown in Fig. 8c. The results are consistent. This result was obtained using only three frames.

4.3. Pool Experiment

We performed a similar experiment at an indoor swimming pool. The upward camera stereo pair was mounted on a tripod and submerged. Several objects resided above the cameras at $Z \in [1, 2]\text{m}$. One object (potato head doll) was placed 1.6m above the rig. The 3D position estimated by our method evolved over time t , as N_{frames} in Eq. (20) increased from 1 to 19. Also the uncertainty of the estimation evolved (gradually reducing). This is illustrated in Fig. 9: the color coding of the ellipsoids indicate temporal evolution of the estimation: blue to red. The estimated height is $\hat{Z} = 1.6 \pm 0.2\text{m}$, consistent with the ground truth. Fig. 10 shows a two dimensional slice of a 3D likelihood function $L(\mathbf{X})$. The slice is a plane defined by $\hat{\mathbf{a}}$ and the WAI points \mathbf{s}^R and \mathbf{s}^L .

In another scenario, two brooms were placed above the cameras at two different positions. Sample frames and the estimated positions are shown in Fig. 11. The result was obtained after 16 frames. Each new frame introduces addi-

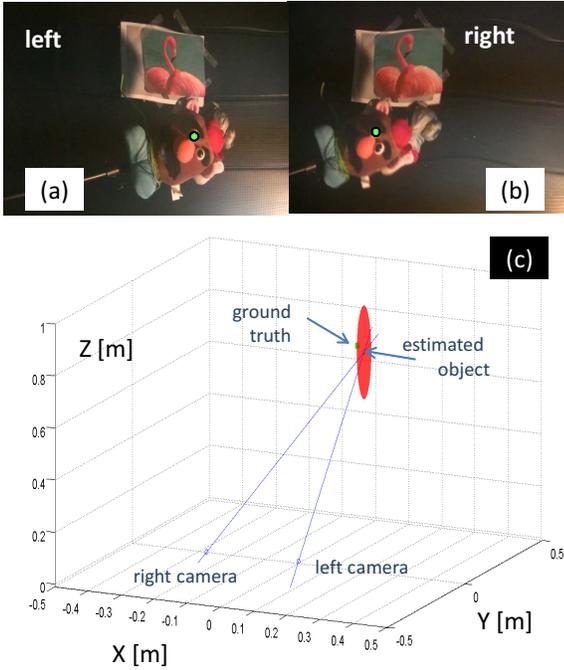


Figure 8. Results of a lab experiment. [a] Sample frame from the left viewpoint. [b] Sample frame from the right viewpoint. [c] The triangulation results in lab coordinates. The ellipsoid represents the uncertainty of the result.

tional information. Thus, the support of the 3D likelihood function shrinks with time. This is shown in Fig. 12.

5. Scaling of Uncertainties

Consider Figs. 8-12. Notice the elongation of the ellipsoids towards the \hat{Z} direction. This section analyzes this phenomenon. In a monocular view, a camera pixel corresponds to an infinite 3D cone which is a region of potential 3D positions. This occurs even in open air without random refractive distortions. The lateral uncertainty ΔX is

$$\Delta X \propto Z \Delta x, \quad (23)$$

where Δx is the pixel size. Viewing monocularly through random refraction distortions, effectively increases the lateral pixel uncertainty Δx , as expressed in Eq. (13) and Fig. 7. This increase in Δx leads to increase of ΔX by Eq. (23).

In stereo, there is uncertainty in depth estimation due to the uncertain disparity. This too, exists without random refractive distortions, as in clear open air. The uncertainty in disparity is proportional to Δx . It can be shown that the depth uncertainty is

$$\Delta Z \propto \frac{Z^2}{b} \Delta x. \quad (24)$$

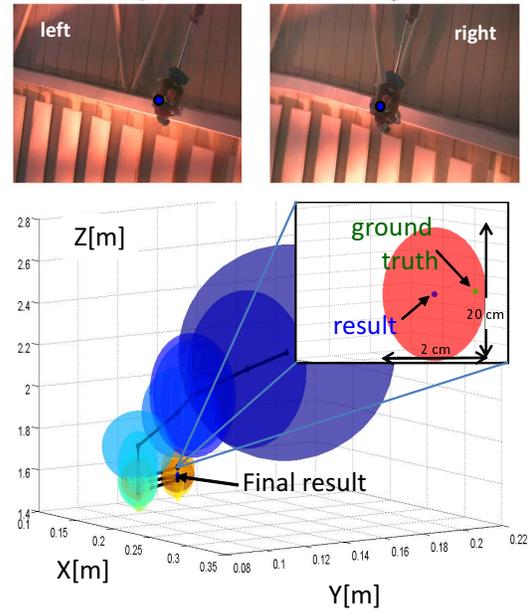


Figure 9. Results of a pool experiment, triangulating a potato head. The illustration shows the estimation evolving as t and N_{frames} increase. Ellipsoids represent the uncertainty of the result, equivalent to the location STD. Time is color coded: blue to red. A zoom-in shows the final result and the ground truth.

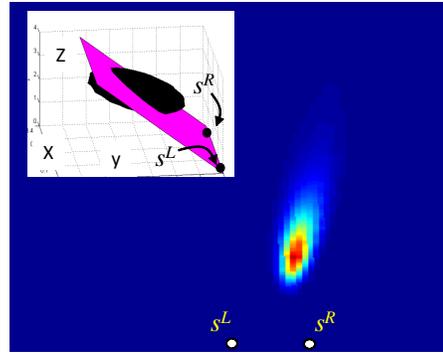


Figure 10. Illustrating the likelihood function in the experiment described in Fig. 9. Shown is $L(\mathbf{X})$ over a plane that includes the water surface points s^R and s^L and the point of ML [insert].

Again, a wavy WAI, effectively increases Δx , and accordingly Eq. (24) increases.

The ratio between lateral and depth uncertainties (Eqs. 23,24) is

$$\frac{\Delta X}{\Delta Z} = \frac{b}{Z}. \quad (25)$$

Since usually $b \ll Z$, it follows that typically $\Delta X \ll \Delta Z$. Thus, the uncertainty ellipsoid is elongated along \hat{Z} .

Note that the relation in Eq. (25) does not depend on

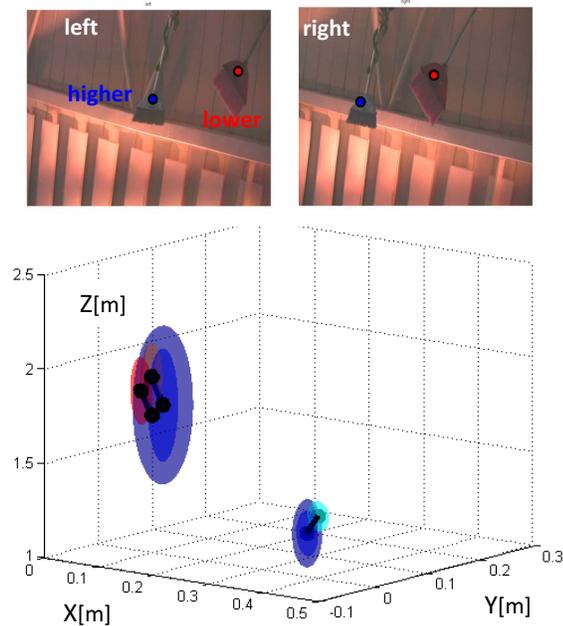


Figure 11. A pool experiment, triangulating brooms. The result is given in lab coordinates. Ellipsoids represent the uncertainty of the result, equivalent to the location STD. Time is color coded: blue to red.

refraction, water waves or our algorithm. It is a general relation that holds for any typical stereo setting, regardless of our settings and method.

6. Discussion

This work deals with triangulation through highly random refractive distortions. The formulation is stochastic, and optimal estimation is obtained in the sense of maximum likelihood. Furthermore, the approach yields the location uncertainty, by the effective support of the 3D likelihood function.

The approach we present here may generalize to detection and analysis of moving objects, through dynamic random refraction. This task was treated in a monocular view [1], but has yet to exploit multiview data. It may be possible also to generalize our approach to lightfield or integral imaging [18, 38, 57], and thus, to acquire many images simultaneously. The method can possibly be used through atmospheric turbulence [24, 50, 58]. There, distortions are typically less severe than through a WAI. However, the operating ranges are much longer.

Acknowledgments

We thank the anonymous reviewers for their useful comments and Joseph Shamir for useful discussions. We thank

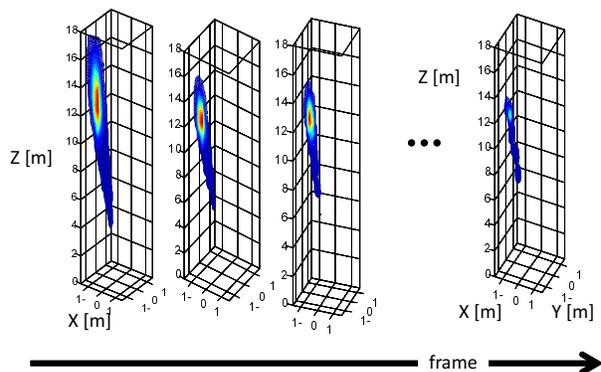


Figure 12. The likelihood function $L(\mathbf{X})$ of a point on the pool ceiling is illustrated by a set of isosurfaces. As data accumulates with time, the support of the likelihood function shrinks, reducing the uncertainty of the estimation.

Daniel Yagodin for his help in the lab experiment, and the swimming pool facilities of the Technion for letting us to conduct the experiments. Yoav Schechner is a Landau Fellow - supported by the Taub Foundation. The work was supported by Israel Science Foundation (ISF) (Grant 1467/12) and the R. L. Kohns Eye Research Fund. This work was conducted in the Ollendorff Minerva Center. Minerva is funded through the BMBF.

References

- [1] M. Alterman, Y. Y. Schechner, J. Shamir, and P. Perona. Detecting motion through dynamic refraction. *IEEE TPAMI*, 35:245–251, and online supplementary material at IEEEExplore, 2013.
- [2] S. Baglio, C. Faraci, and E. Foti. Structured light approach for measuring sea ripple characteristics. In *IEEE OCEANS*, volume 1, pages 449–453, 1998.
- [3] A. Ben-Simon, O. Ben-Shahar, G. Vasserman, M. Ben-Tov, and R. Segev. Visual acuity in the archerfish: Behavior, anatomy, and neurophysiology. *Journal of Vision*, 12(12), 2012.
- [4] J. Y. Bouguet. Camera calibration toolbox for matlab, www.vision.caltech.edu/bouguetj/calib_doc.
- [5] T. Brox. Large displacement optical flow code. <http://132.230.167.110/people/brox/resources/pami2010Matlab.zip>.
- [6] T. Brox and J. Malik. Large displacement optical flow: Descriptor matching in variational motion estimation. *IEEE TPAMI*, 33:500–513, 2011.
- [7] Y. J. Chang and T. Chen. Multi-view 3D reconstruction for scenes under the refractive plane with known vertical direction. *IEEE ICCV*, 2011.
- [8] V. Chari and P. Sturm. Multi-view geometry of the refractive plane. In *Proc. BMVC*, 2009.
- [9] C. Cox and W. Munk. Statistics of the sea surface derived from sun glitter. *J. Mar. Res.*, 13:198–227, 1954.

- [10] D. Dabiri and M. Gharib. Interaction of a shear layer with a free surface. In *Int. Sympos. Flow Vis.*, 2000.
- [11] F. Dellaert, S. Seitz, C. Thorpe, and S. Thrun. Structure from motion without correspondence. In *IEEE CVPR*, volume 2, pages 557–564, 2000.
- [12] Y. Ding, F. Li, Y. Ji, and J. Yu. Dynamic 3D fluid surface acquisition using a camera array. *IEEE ICCV*, 2011.
- [13] L. S. Dolin, A. G. Luchinin, V. I. Titov, and D. G. Turlaev. Correcting images of underwater objects distorted by sea surface roughness. *Proc. of SPIE*, 6615, 2007.
- [14] A. Donate and E. Ribeiro. Improved reconstruction of images distorted by water waves. *Advances in Computer Graphics and Computer Vision*, pages 264–277, 2007.
- [15] A. Efros, V. Isler, J. Shi, and M. Visontai. Seeing through water. *NIPS*, 17:393–400, 2004.
- [16] R. Ferreira, J. Costeira, and J. Santos. Stereo reconstruction of a submerged scene. *Pattern Recognition and Image Analysis*, pages 102–109, 2005.
- [17] M. Gupta, S. G. Narasimhan, and Y. Y. Schechner. On controlling light transport in poor visibility environments. In *IEEE CVPR*, pages 1–8, 2008.
- [18] S. H. Hong, J. S. Jang, and B. Javidi. Three-dimensional volumetric object reconstruction using computational integral imaging. *Optics Express*, 12(3):483–491, 2004.
- [19] G. Horváth and D. Varjú. On the structure of the aerial visual field of aquatic animals distorted by refraction. *Bull. Math. Bio.*, 53(3):425–441, 1991.
- [20] I. Ihrke, K. Kutulakos, H. P. Lensch, M. A. Magnor, and W. Heidrich. State of the art in transparent and specular object reconstruction. *Proc. Eurographics*, 2008.
- [21] K. Ikeuchi, M. Sakauchi, H. Kawasaki, and I. Sato. Constructing virtual cities by using panoramic images. *IJCV*, 58(3):237–247, 2004.
- [22] B. Jähne, J. Klinke, and S. Waas. Imaging of short ocean wind waves: a critical theoretical review. *JOSA*, 11:2197–2209, 1994.
- [23] A. Jorđt-Sedlazeck and R. Koch. Refractive calibration of underwater cameras. In *Proc. ECCV*, pages 846–859, 2012.
- [24] N. Joshi and M. Cohen. Seeing mt. rainier: Lucky imaging for multi-image denoising, sharpening, and haze removal. In *IEEE ICCP*, 2010.
- [25] G. Katzir and N. Intrator. Striking of underwater prey by a reef heron, *egretta gularis schistacea*. *J. Compar. Phys. A.*, 160:517–523, 1987.
- [26] M. J. Kidger. *Fundamental Optical Design*. SPIE Press, 2002.
- [27] H. Kleerekoper and T. Malar. Orientation through sound in fishes in hearing mechanisms in vertebrates, 1968.
- [28] S. Koppal, I. Gkioulekas, T. Zickler, and G. Barrows. Wide-angle micro sensors for vision on a tight budget. In *IEEE CVPR*, pages 361–368, 2011.
- [29] M. Loktev, O. Soloviev, S. Savenko, and G. Vdovin. Speckle imaging through turbulent atmosphere based on adaptable pupil segmentation. *Optics Letters*, 36(14):2656–2658, 2011.
- [30] D. M. Milder, P. W. Carter, N. L. Flacco, B. E. Hubbard, N. M. Jones, K. R. Panici, B. D. Platt, R. E. Potter, K. W. Tong, and D. J. Twisselmann. Reconstruction of through-surface underwater imagery. *Waves in Random and Complex Media*, 16:521 – 530, 2006.
- [31] D. Miyazaki, M. Saito, Y. Sato, and K. Ikeuchi. Determining surface orientations of transparent objects based on polarization degrees in visible and infrared wavelengths. *JOSA A*, 19(4):687–694, 2002.
- [32] A. Molkov and L. Dolin. Determination of wind roughness characteristics based on an underwater image of the sea surface. *Izvestiya, Atmospheric and Oceanic Physics*, 48:552–564, 2012.
- [33] N. Morris and K. Kutulakos. Dynamic refraction stereo. In *IEEE ICCV*, volume 2, pages 1573–1580, 2005.
- [34] H. Murase. Surface shape reconstruction of a nonrigid transport object using refraction and motion. *IEEE TPAMI*, 14:1045–1052, 1992.
- [35] O. Oreifej, G. Shu, T. Pace, and M. Shah. A two-stage reconstruction approach for seeing through water. *IEEE CVPR*, 2011.
- [36] S. Oron, A. Bar-Hillel, D. Levi, and S. Avidan. Locally orderless tracking. In *IEEE CVPR*, pages 1940–1947, 2012.
- [37] S. Peleg, M. Ben-Ezra, and Y. Pritch. Omnistereo: Panoramic stereo imaging. *IEEE CVPR*, 23(3):279–290, 2001.
- [38] Raytrix. <http://www.raytrix.de>.
- [39] H. Saito, H. Kawamura, and M. Nakajima. 3D shape measurement of underwater objects using motion stereo. In *IEEE IECON*, volume 2, pages 1231–1235, 2002.
- [40] Y. Y. Schechner and N. Karpel. Attenuating natural flicker patterns. In *MTS/IEEE Oceans*, pages 1262–1268, 2004.
- [41] M. S. Schmalz. Integration of stereophotogrammetry with image restoration models for distortion-tolerant viewing through the sea surface. *Proc. SPIE*, 1943:115–128, 1993.
- [42] H. Schultz. Shape reconstruction from multiple images of the ocean surface. *Surface, Photogrammetric Engineering and Remote Sensing*, 62:93–99, 1994.
- [43] H. Schultz and A. Corrada-Emmanuel. System and method for imaging through an irregular water surface, 2007. US Patent 7,630,077.
- [44] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: Exploring image collections in 3D. In *Proc. SIGGRAPH*, 2006.
- [45] J. B. Steen. The swim bladder as a hydrostatic organ. *Fish Physiology: The nervous system, circulation, and respiration*, 4, 1970.
- [46] H. Suiter, N. Flacco, P. Carter, K. Tong, R. Ries, and M. Gershenson. Optics Near the Snell Angle in a Water-to-Air Change of Medium. In *IEEE OCEANS*, pages 1–12, 2008.
- [47] Y. Swirski and Y. Y. Schechner. 3Deflicker from motion. In *IEEE ICCP*, 2013.
- [48] Y. Swirski, Y. Y. Schechner, B. Herzberg, and S. Negahdaripour. CauStereo: Range from light in nature. *App. Opt.*, 50:89–101, 2011.
- [49] Y. Swirski, Y. Y. Schechner, and T. Nir. Variational stereo in dynamic illumination. In *IEEE ICCV*, 2011.
- [50] Y. Tian, S. Narasimhan, and A. Vannevel. Depth from optical turbulence. *IEEE CVPR*, 2012.

- [51] Y. Tian and S. G. Narasimhan. Seeing through water: Image restoration using model-based tracking. In *IEEE ICCV*, pages 2303–2310, 2009.
- [52] Y. Tian and S. G. Narasimhan. A globally optimal data-driven approach for image distortion estimation. *IEEE CVPR*, 2010.
- [53] T. Treibitz, Y. Y. Schechner, and H. Singh. Flat refractive geometry. *IEEE CVPR*, 2008.
- [54] Y. Tsin, S. Kang, and R. Szeliski. Stereo matching with reflections and translucency. In *IEEE CVPR*, volume 1, pages I–702, 2003.
- [55] Z. Wen, A. Lambert, D. Fraser, and H. Li. Bispectral analysis and recovery of images distorted by a moving water surface. In *JOSA A*. OSA, 2010.
- [56] R. Westaway, S. Lane, and D. Hicks. Remote sensing of clear-water, shallow, gravel-bed rivers using digital photogrammetry. *Photogrammetric Engineering and Remote Sensing*, 67(11):1271–1282, 2001.
- [57] G. Wetzstein, D. Roodnick, W. Heidrich, and R. Raskar. Refractive shape from light field distortion. In *IEEE ICCV*, pages 1180–1186, 2011.
- [58] X. Zhu and P. Milanfar. Removing atmospheric turbulence via space-invariant deconvolution. *IEEE TPAMI*, 35(1):157–170, 2013.