# Triangulation in Random Refractive Distortions

Marina Alterman, Yoav Y. Schechner, *Member, IEEE*, and Yohay Swirski

**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.

**Index Terms**—Underwater, stereo, triangulation, probability, likelihood

## 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. The biological world does not necessarily indicate solutions we need to take, but it motivates the definition of the problem. As illustrated in Fig. 2a, birds [1] fly in search of submerged fish to hunt. Some animals hunt the other way. For example, the archer fish [2] launches water jets in a parabolic ballistic trajectory to shoot down flies (Fig. 2b) that sit on foliage. In all cases, predators need to well assess the three dimensional (3D) location of the prey prior to charging. This can be done by triangulation as the predator changes its position during its path, thus obtaining multiple views. Additional biological details and motivations appear in [3]. Upward vision through a wavy WAI can serve as a *virtual periscope* (Fig. 2c). This can help submarines assess activities above water, without using physical periscopes which flag their presence.

In the open air, triangulation leading to 3D scene reconstruction is well studied [4], [5], [6], [7], [8], [9]. Distortions through a *flat* WAI have also been thoroughly studied [10], [11], [12], [13], [14], [15], [16], [17]. Measuring from air the 3D structure of submerged scenes has been proposed based on stereo [12] or motion [10], [16]. Still, a triangulation challenge remains when a flat-WAI is perturbed by random unknown WAI waves. This paper addresses this challenge. Furthermore, this paper shows that *multiplying the viewpoints 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.

We take a stochastic approach to triangulation under random refractive distortions. We acquire a video sequence of an airborne object, using an underwater camera stereo pair. The statistics of WAI waves induce by refraction a probabilistic relation between any pixel location and an airborne LOS. Inspired by [18], measurements of an object's random projection from multiple views and times lead to a likelihood function of the object's 3D location. Maximum likelihood (ML) estimates the 3D location, while the effective support of the likelihood function informs of the location uncertainty.

In addition, we formulate the 3D ML problem as a minimization based directly on image plane coordinates. This simplifies the computations and reduces run time. We use the method from [19] to obtain multiple point correspondences between views. Preliminary results and theory appeared in [20].

## 2 Theoretical Background

Refraction has been used and analyzed in the context of computational photography [21], [22], [23], [24], [25], [26], [27]. Theory about visual refraction through a WAI is described in Ref. [28]. This section briefly follows notations and relevant derivations from Ref. [28].

### 2.1 Snell's Law in Various Forms

Consider Fig. 3. Let us back-project a LOS from a submerged camera towards an airborne object point. The LOS

- *M. Alterman is with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208. E-mail: amarinago@gmail.com.*
- *Y.Y. Schechner and Y. Swirski are with the Viterbi Faculty of Electrical Engineering, Technion, Israel Institute of Technology, Haifa 32000, Israel. E-mail: syohays@gmail.com, yoav@ee.technion.ac.il.*

*Manuscript received 18 Feb. 2015; revised 14 Mar. 2016; accepted 17 Mar. 2016. Date of publication 6 Apr. 2016; date of current version 13 Feb. 2017. Recommended for acceptance by P. Sturm.*
*For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.*
*Digital Object Identifier no. 10.1109/TPAMI.2016.2551740*

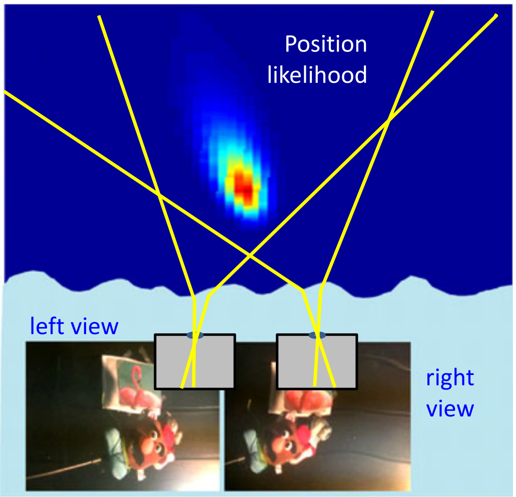

Fig. 1. Images of a sample stereo pair taken by looking upward through a wavy water surface. By triangulation, using our method we obtain a likelihood function in 3D. The estimated location is the position of the maximum likelihood (ML).

in water is directed along unit vector $\hat{\mathbf{v}}_{\mathrm{w}}$. This vector forms an angle $\theta_{\mathrm{w}}$ 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 $\hat{\mathbf{v}}_{\mathrm{a}}$. This vector forms an angle $\theta_{\mathrm{a}}$ with $\hat{\mathbf{N}}$. The scalar Snell's law of refraction is

$$\sin\theta_{\mathrm{a}} = n\sin\theta_{\mathrm{w}} \ , \tag{1}$$

i.e,

$$\cos\theta_{\mathrm{a}} = \sqrt{1 - n^2 + n^2\cos^2\theta_{\mathrm{w}}}, \tag{2}$$

where $n \approx 1.33$ is the optical refractive index of water. Vector forms of Snell's law [29] are

$$\hat{\mathbf{v}}_{\mathrm{a}} \times \hat{\mathbf{N}} = n\hat{\mathbf{v}}_{\mathrm{w}} \times \hat{\mathbf{N}}. \tag{3}$$

$$\hat{\mathbf{v}}_{\mathrm{a}} = n\hat{\mathbf{v}}_{\mathrm{w}} + \hat{\mathbf{N}}(\cos\theta_{\mathrm{a}} - n\cos\theta_{\mathrm{w}}), \tag{4}$$

where

$$\cos\theta_{\mathrm{w}} = \hat{\mathbf{v}}_{\mathrm{w}} \cdot \hat{\mathbf{N}}. \tag{5}$$

Here $\times$ is the cross product. From Eq. (3), $\hat{\mathbf{v}}_{\mathrm{w}}$, $\hat{\mathbf{v}}_{\mathrm{a}}$ 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 (not shown). Its origin is at the center of projection $\mathbf{O}$. The axes include the optical axis and the lateral pixel coordinates of the image plane. A pixel length is $h_{\mathrm{pixel}}$. The optical axis intersects the image plane at pixel $\mathbf{c}$. In the camera coordinate system, pixel $\mathbf{x}$ is at physical location

$$\mathbf{X}_{\mathrm{cam}} = \begin{bmatrix} \mathbf{x} - \mathbf{c} \\ f_{\mathrm{c}} \end{bmatrix} h_{\mathrm{pixel}}, \tag{6}$$

where $f_c$ is the focal length of the camera. The values of $\mathbf{x}, \mathbf{c}$ and $f_c$ are given 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 axes, $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$. In the global coordinate system, the 3D location of pixel $\mathbf{x}$ is

$$\mathbf{X}_{\mathrm{lab}} = \mathbf{R}^T \mathbf{x}_{\mathrm{cam}}, \tag{7}$$

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

$$\hat{\mathbf{v}}_{\mathrm{w}} = \mathbf{X}_{\mathrm{lab}} / \|\mathbf{X}_{\mathrm{lab}}\|. \tag{8}$$

Underwater cameras often have a flat-glass interface, due to which the camera-in-housing system has no single center of projection. Then, the relation between $\hat{\mathbf{v}}_{\mathrm{w}}$ and pixel $\mathbf{x}$ is not as simple as Eqs. (6 - 8). Nevertheless, if needed, a comprehensive calibration process [10] deterministically establishes the LOS and thus $\hat{\mathbf{v}}_{\mathrm{w}}$, per $\mathbf{x}$. This paper deals with the stochastic, unpredictable effects of waves. Thus, deterministic calibration matters are assumed done.

We 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

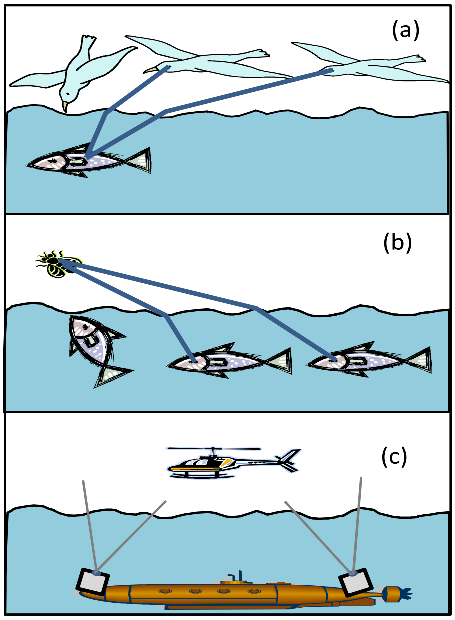

Fig. 2. Relevant scenarios. (a) Birds fly by to hunt for a submerged fish. (b) The archer fish shoots water jets to shoot down airborne flies. (c) A submarine avoiding use of a physical periscope by upward vision through a wavy WAI.

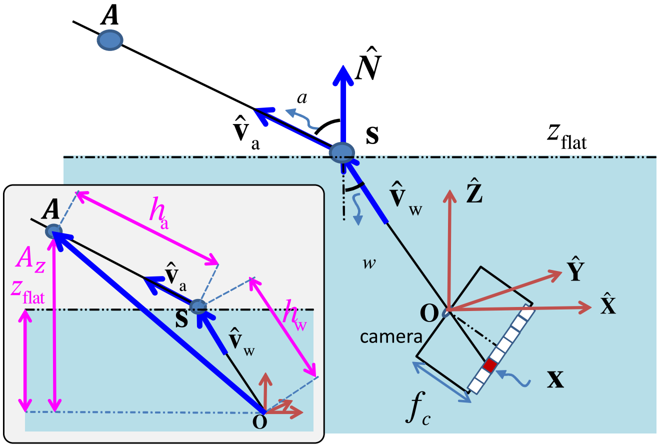

Fig. 3. Viewing geometry through a flat water surface. [Inset] Definition of additional length parameters (magenta).

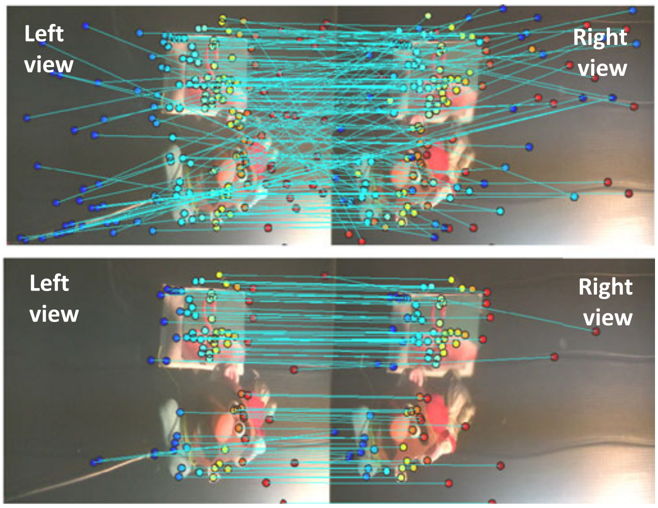

Fig. 4. Colors indicate matching. [Top] Correspondence produced by SIFT [32]. [Bottom] Correspondence produced by [19].

$$\hat{\mathbf{v}}_{\rm a} = n\hat{\mathbf{v}}_{\rm w} + \hat{\mathbf{Z}}\left[\sqrt{1-n^2+n^2(\hat{\mathbf{v}}_{\rm w}\cdot\hat{\mathbf{Z}})^2} - n\hat{\mathbf{v}}_{\rm w}\cdot\hat{\mathbf{Z}}\right]. \qquad (9)$$

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

An airborne object is at $\mathbf{A} = (A_{\rm x}, A_{\rm y}, A_z)^T$ in the global coordinate system (Fig. 3[Inset]). Back-propagating from the camera's center of projection, a refracted LOS includes a submerged LOS segment and an airborne one. Thus,

$$\mathbf{A} = \hat{\mathbf{v}}_{\rm w}h_{\rm w} + \hat{\mathbf{v}}_{\rm a}h_{\rm a} \;\Rightarrow\; A_z = h_{\rm w}v_{{\rm w},z} + h_{\rm a}v_{{\rm a},z}, \qquad (10)$$

where $h_{\rm w}$ and $h_{\rm a}$ are line-length parameters. Suppose the camera is at depth $Z_{\rm flat}$ below a flat WAI. Then, substituting $h_{\rm w}v_{{\rm w},z} = Z_{\rm flat}$ into Eq. (10) yields.

$$h_{\rm w} = Z_{\rm flat}/v_{{\rm w},z}, \quad h_{\rm a} = (A_z - Z_{\rm flat})/v_{{\rm a},z}. \qquad (11)$$

The parameters $Z_{\rm flat}$ and $\mathbf{R}$ can be, in principle, measured by the system. A depth gauge in the system can measure $Z_{\rm flat}$. 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 [10]. In nature, fish can control their balance using otoliths [30] or a swim-bladder [31]. Even without an accelerometer, the inclination angle can be computed by the location of Snell's window in the field of view [28].

### 2.3 Image Correspondence

Consider a submerged stereo system having a baseline $b$. Variables associated with the left or right camera are denoted by L or R, respectively. The projection of an object point $\mathbf{a}$ through camera L is $\mathbf{x}^{\rm L}$. The projection of $\mathbf{a}$ through camera R is $\mathbf{x}^{\rm R}$. The image pair, taken by the left and right cameras, is given as input. Correspondence then needs to be established between $\mathbf{x}^{\rm L}$ and $\mathbf{x}^{\rm R}$ for various object points. Establishing correspondence is a fundamental problem in computer vision. Finding corresponding points in our scenario is difficult. The reason is that the shape of objects and their appearance can change across images, particularly due to the randomness of refractive distortions.

The SIFT algorithm [32], for example, alongside correct correspondences, typically produces a considerable number of incorrectly matched outliers especially in refractive distortions (see Fig. 4[Top]). To make these correspondences useful, outliers must be removed. We use the method in [19] to filter these outliers (see Fig. 4 [Bottom]). It uses geometric consistency to separate outliers from correct matches. This is done by modeling bounded distortion between images. Bounded distortion accounts for some deformation while preserving the geometric information that enables correct correspondences. The algorithm solves an optimization problem which results in a maximal set of corresponding points $\{(\mathbf{x}^{\rm L}_m, \mathbf{x}^{\rm R}_m)\}_{m=1}^{N_{\rm matches}}$. We then use Large Displacement Optical Flow (LDOF) [33], [34] for tracking the initially matched point-set through temporal frames, to obtain temporal correspondence.

## 3 Modeling Position Statistics

### 3.1 Single View Random Projection Distribution

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

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

Here $\mathbf{d}$ is the displacement in the image of the object, caused by the random WAI waves. In other words, the spatiotemporal field $\mathbf{d}(\mathbf{x}_{\rm flat}, t)$ is a *random distortion* created by random refractive changes. Fig. 5[Left] illustrates the distorted pixel positions $\mathbf{x}$ around $\mathbf{x}_{\rm flat}$. Following Cox and Munk [35], the WAI normal $\hat{\mathbf{N}}$ is random in space and time and has a Gaussian distribution. The variance of $\hat{\mathbf{N}}$ depends on meteorological parameters. For a given $\mathbf{x}_{\rm flat}$, the random vector $\mathbf{d}$ has approximately a normal [28] distribution: $\mathbf{d} \sim \mathcal{N}(0, \Sigma_{\mathbf{x}})$. Thus, the probability density function (PDF) of imaging $\mathbf{a}$ at $\mathbf{x}$ is approximated by

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

where $G$ is a normalization factor.

The $2\times 2$ covariance matrix $\Sigma_{\mathbf{x}}$ depends on the WAI roughness, the camera parameters $(\mathbf{R}, h_{\rm pixel}, f_c)$ and somewhat on $\mathbf{x}_{\rm flat}$. As described in [28], physics-based simulations can derive an approximate yet useful $\Sigma_{\mathbf{x}}$ without empirical image data. Empirically, prior to triangulation attempts, the statistics of distortion can be learned [28]. Object points known to be static, can be observed for a short while, e.g., a few seconds, providing samples of $\mathbf{x} \sim \mathcal{N}(\mathbf{x}_{\rm flat}, \Sigma_{\mathbf{x}})$. A Gaussian can be fitted to the tracked projections, over several frames, thus empirically estimating $\Sigma_{\mathbf{x}}$. This process is analogous to [36], where a laser beam is reflected by a WAI onto a screen, demonstrating a point spread function. Once $\Sigma_{\mathbf{x}}$ is set, it is used later to stochastically triangulate objects at an instant, without a necessity to accumulate video data.

Fig. 5[Middle] shows the Gaussian fitted to the distribution in Fig. 5[Left]. The mean of the points in Fig. 5 [Left] estimates $\mathbf{x}_{\rm flat}$. Thus Fig. 5[Middle] approximately illustrates the probability of $\mathbf{x}$ given $\mathbf{x}_{\rm flat}$. In the task of

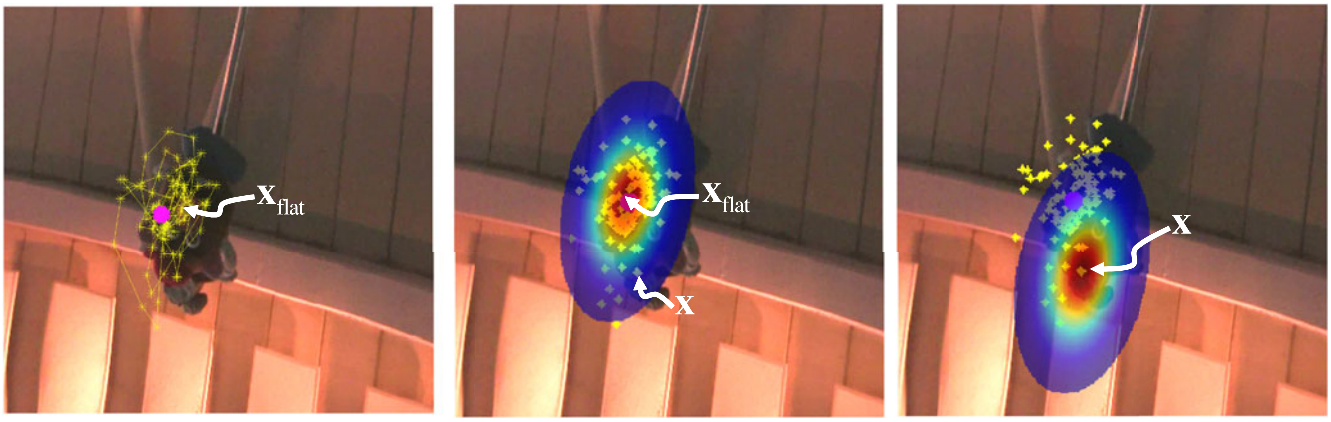

Fig. 5. [Left] All the pixels $\mathbf{x}$ corresponding to an object point over time. The mean location estimates $\mathbf{x}_{\text{flat}}$. [Middle] The distribution of $\mathbf{x}$ given $\mathbf{x}_{\text{flat}}$. The distribution of possible pixel positions is approximately normal around $\mathbf{x}_{\text{flat}}$. Red expresses high probability and blue low probability. [Right] The distribution of $\mathbf{x}_{\text{flat}}$ given $\mathbf{x}$.

this work, $\mathbf{x}_{\text{flat}}$ is not given, and there may only be a single temporal instance in which $\mathbf{x}$ is measured. According to Bayes' rule, the probability of $\mathbf{x}_{\text{flat}}$ given $\mathbf{x}$ is,

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

The PDF $p(\mathbf{x}_{\text{flat}})$ is a prior on where the object might preferably be projected to, when the WAI is flat. If there is a prior on the object's projection, it can be incorporated. Often, there is no preferred object location. Then, $p(\mathbf{x}_{\text{flat}})$ is a constant, and

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

Based on Eq. (15), the probability $p(\mathbf{x}|\mathbf{x}_{\text{flat}})$ in Fig. 5[Middle] is displaced to be centered at $\mathbf{x}$. Thus, Fig. 5[Right] illustrates the probability $p(\mathbf{x}_{\text{flat}}|\mathbf{x})$.

### 3.2 Single View Airborne Position Likelihood

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

Consider Fig. 6. Since the WAI is wavy, $\mathbf{A}$ projects to pixel $\mathbf{x}^{\text{L}}(t)$, at time $t$, while $\mathbf{x}^{\text{L}}_{\text{flat}}$ is unknown. Eq. (15) sets a probability density $p[\mathbf{x}^{\text{L}}_{\text{flat}}|\mathbf{x}^{\text{L}}(t)]$ to each pixel $\mathbf{x}^{\text{L}}_{\text{flat}}$ in the L image plane. Thus, Eqs. (13) and (15) set a probability density

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

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

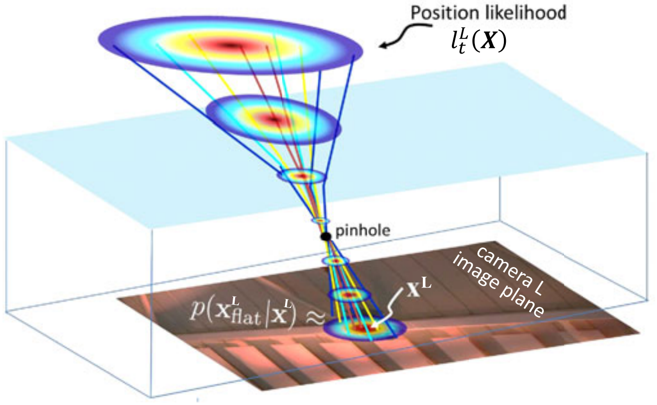

Fig. 6. Position likelihood. Back-projecting all the pixels around $\mathbf{x}$ based on $p(\mathbf{x}_{flat}|\mathbf{x})$ through a flat surface yields a 3D cone. This cone represents the position likelihood of the object imaged at $\mathbf{x}$.

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 $\mathbf{X} \in \text{LOS}(\mathbf{x}^{\text{L}}_{\text{flat}})$, we associate a *likelihood* equivalent to the probability density defined in Eq. (16):

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

Based on Eqs. (16) and (17)

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

Note that all $\mathbf{X}$ that project to the same $\mathbf{x}^{\text{L}}_{\text{flat}}$ have the same likelihood,[1] given a measured location $\mathbf{x}^{\text{L}}(t)$. Thus a probability in 2D induces a score $l^{\text{L}}_t(\mathbf{X})$ in 3D.

Suppose that distortions between frames are statistically independent. Then, likelihoods stemming from different temporal frames multiply each other. Overall, for $N_{\text{frames}}$ frames, the likelihood is

$$L(\mathbf{X}) = \prod_{t=1}^{N_{\text{frames}}} l^{\text{L}}_t(\mathbf{X}). \tag{19}$$

### 3.3 Uncorrelated Multi-Views

We now study random refraction in multiview geometry. Each camera views the object through a different WAI portion: one is around horizontal location $\mathbf{S}^{\text{L}} = (S^{\text{L}}_x, S^{\text{L}}_y)$, while the other is around $\mathbf{S}^{\text{R}} = (S^{\text{R}}_x, S^{\text{R}}_y)$. The projected image point on the left is distorted (displaced) by $\mathbf{d}^{\text{L}}$, while the corresponding point on the right image is displaced by $\mathbf{d}^{\text{R}}$.

[1] This likelihood score does not integrate to 1 over an unbounded 3D domain. It is possible to normalize this score in an artificially bounded domain, but such normalization does not change the later optimization and is hence skipped.

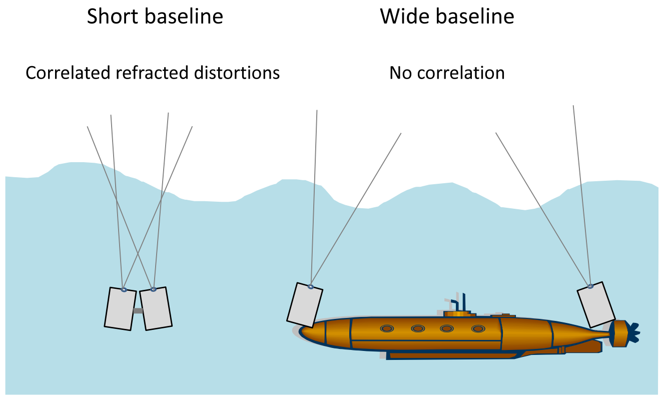

Fig. 7. [Left] A baseline that is significantly shorter than the typical correlation length of the WAI-slope, results in highly correlated distortions across views. [Right] A wide baseline results in uncorrelated distortions between the two views.

Is the random dynamic distortion $\mathbf{d}^{\mathrm{L}}(t)$ similar or very different than $\mathbf{d}^{\mathrm{R}}(t), \forall t$? How mutually statistically dependent are $\mathbf{d}^{\mathrm{L}}$ and $\mathbf{d}^{\mathrm{R}}$? The distortion mainly depends on the WAI slope [28], i.e., on $\hat{\mathbf{N}}(\mathbf{S}, t)$. Hence, statistical dependency between $\mathbf{d}^{\mathrm{L}}$ and $\mathbf{d}^{\mathrm{R}}$ is dictated by the dependency between $\hat{\mathbf{N}}(\mathbf{S}^{\mathrm{L}})$ and $\hat{\mathbf{N}}(\mathbf{S}^{\mathrm{R}})$. According to Ref. [35], the WAI slope statistics is approximately Gaussian. Hence, statistical dependency of slopes is equivalent to correlation. The correlation between slopes[2] in $\mathbf{S}^{\mathrm{L}}$ and $\mathbf{S}^{\mathrm{R}}$ dictates the correlation between $\mathbf{d}^{\mathrm{L}}$ and $\mathbf{d}^{\mathrm{R}}$.

Correlation between WAI slopes at $\mathbf{S}^{\mathrm{L}}$ and $\mathbf{S}^{\mathrm{R}}$ depends on the lateral distance $b_s \approx \|\mathbf{S}^{\mathrm{L}} - \mathbf{S}^{\mathrm{R}}\|$. If the baseline $b_s$ is significantly shorter than the typical WAI-slope *correlation length*, then distortions in the two views are mutually highly correlated, as illustrated in Fig. 7. We prefer to work in a domain, where distortions in the multiple views are mutually *uncorrelated*, largely. Low statistical dependency (low mutual information) between multiview measurements means that any new view adds *more information* about the object location. This is achieved if $b_s$ is significantly larger than the typical WAI-slope correlation length (See Fig. 7). We do not know the WAI-slope correlation length a-priori. However, in our experiments, we verified empirically the low correlation between $\mathbf{d}^{\mathrm{L}}$ and $\mathbf{d}^{\mathrm{R}}$, due to commonly occurring small WAI wiggles.

## 4 Multi-View Stochastic Estimation

Section 3.2 defined variables associated with the L camera. Let us generalize the formulation to multiple views. Under a flat WAI, the object in $\mathbf{A}$ projects to pixel $\mathbf{x}^{\mathrm{R}}_{\mathrm{flat}}$ in camera R. Since the WAI is wavy, $\mathbf{A}$ projects to pixel $\mathbf{x}^{\mathrm{R}}(t)$, at time $t$, while $\mathbf{x}^{\mathrm{R}}_{\mathrm{flat}}$ is unknown. Similarly to the process involving Eq. (18), we derive the likelihood $l^{\mathrm{R}}_t(\mathbf{X}), \forall \mathbf{X}$, based on

$$l^{\mathrm{R}}_t(\mathbf{X}) \sim p\left[\mathbf{x}^{\mathrm{R}}(t)|\mathbf{x}^{\mathrm{R}}_{\mathrm{flat}}\right] \quad | \ \mathbf{X} \in \mathrm{LOS}\left(\mathbf{x}^{\mathrm{R}}_{\mathrm{flat}}\right). \tag{20}$$

Following Section 2.3, suppose that between views, correspondence of image points is established. In other words, we know that the measurement pixel set $\{\mathbf{x}^{\mathrm{L}}(t), \mathbf{x}^{\mathrm{R}}(t)\}_t$ corresponds to the same 3D object point, but we do not know where the object is.

From Section 3.3, distortions between views are approximated as statistically independent. Therefore, likelihoods stemming from different frames and viewpoints multiply each other. Overall, for $N_{\mathrm{frames}}$ frames, the likelihood (19) generalizes to

$$L(\mathbf{X}) = \prod_{t=1}^{N_{\mathrm{frames}}} l^{\mathrm{L}}_t(\mathbf{X}) l^{\mathrm{R}}_t(\mathbf{X}), \tag{21}$$

as illustrated in Fig. 8.

For $N_{\mathrm{views}}$ views, Eq. (21) generalizes to

$$L(\mathbf{X}) = \prod_{t=1}^{N_{\mathrm{frames}}} \prod_{i=1}^{N_{\mathrm{views}}} l^{(i)}_t(\mathbf{X}) \quad , \tag{22}$$

where $(i)$ is the view index. In a stereo camera pair $(i) \in \{\mathtt{L}, \mathtt{R}\}$. The effective 3D spatial support (orange region in Fig. 8) 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. (22) using multiple likelihood factors. Hence, increasing the number of viewpoints can *shorten* the acquisition time needed for certain localization accuracy of the object (see Fig. 9). ML yields an estimate of an optimal airborne object location.

$$\hat{\mathbf{A}} = \arg\max_{\mathbf{X}} L(\mathbf{X}). \tag{23}$$

We summarize the method in Algorithm 1.

**Algorithm 1.** Estimation Using a 3D Likelihood Function

```
Data: $N_{\mathrm{frames}}$ distorted stereo images of an object at $\mathbf{A}$.
Result: Estimated position $\hat{\mathbf{A}}$ of an object at $\mathbf{A}$.
$L = 1$.
for $t = 1$ to$N_{\mathrm{frames}}$ do
    if $t = 1$ then
        Select pixels $\mathbf{x}^{\mathrm{L}}(1)$ and $\mathbf{x}^{\mathrm{R}}(1)$ which correspond to $\mathbf{A}$.
    else
        Track the object from time $t-1$ to time $t$ to get $\mathbf{x}^{\mathrm{L}}(t)$ and $\mathbf{x}^{\mathrm{R}}(t)$.
    end
    Back-project $p[\mathbf{x}^{\mathrm{L}}_{\mathrm{flat}}|\mathbf{x}^{\mathrm{L}}(t)]$ and $p[\mathbf{x}^{\mathrm{R}}_{\mathrm{flat}}|\mathbf{x}^{\mathrm{R}}(t)]$ through a flat WAI to get $l^{\mathrm{L}}_t(\mathbf{X})$ and $l^{\mathrm{R}}_t(\mathbf{X})$.
    $L(\mathbf{X}) \leftarrow L(\mathbf{X}) l^{\mathrm{L}}_t(\mathbf{X}) l^{\mathrm{R}}_t(\mathbf{X})$.
end
$\hat{\mathbf{A}} = \arg\max_{\mathbf{X}} L(\mathbf{X})$.
return $\hat{\mathbf{A}}$.
```

We work with discrete, regular spatial grids: image locations $\mathbf{x}^{\mathrm{L}}(t)$, $\mathbf{x}^{\mathrm{L}}_{\mathrm{flat}}$ 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 $l^{\mathrm{L}}_t(\mathbf{X}), \forall \mathbf{X}$.

Estimation based on random noisy data has an experimental uncertainty. There are various ways to state

[2] Ref. [37] proposes a model to characterize WAI slope and curvature variances, as well as their spatial correlation.

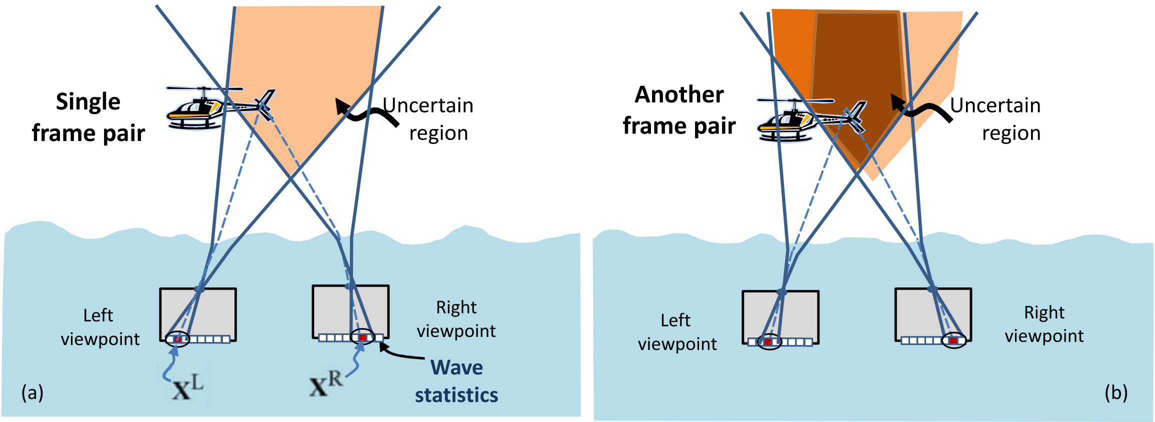

Fig. 8. (a) By projecting the uncertainties around $\mathbf{x}^{\mathrm{L}}$ and $\mathbf{x}^{\mathrm{R}}$, the effective 3D spatial support (orange region) of the overlap represents the 3D domain in which the airborne object point is likely to reside. (b) Additional temporal frames narrow this domain.

uncertainty, including confidence intervals or standard deviation (STD). One simple way is to threshold the likelihood, to determine a set $\Psi$ of object locations

$$\Psi = \{\mathbf{X} : \ [L(\mathbf{X})/L(\hat{\mathbf{A}})] > \tau\}, \tag{24}$$

whose likelihood is significant. Here $\tau$ is a threshold. The spatial bounds of $\Psi$ set an uncertainty assessment. We select the minimum and maximum coordinates of the elements of $\Psi$ to determine an axis-aligned spatial bounding box.

The estimation described above uses a 3D likelihood function. Next, we present another method, using probabilities in 2D images.

### Position Estimation from Probabilities in Images

The method presented in Section 4 involves a 3D likelihood function. In some cases, it is possible to simplify the estimation, using probabilities in the 2D images. The likelihood of $\mathbf{X}$ is

$$L(\mathbf{X}) = \prod_{t=1}^{N_{\text{frames}}} p[\mathbf{x}^{\mathrm{L}}_{\text{flat}}(\mathbf{X})|\mathbf{x}^{\mathrm{L}}(t)]p[\mathbf{x}^{\mathrm{R}}_{\text{flat}}(\mathbf{X})|\mathbf{x}^{\mathrm{R}}(t)]. \tag{25}$$

For simplicity of computations, instead of computing the maximum of $L(\mathbf{X})$ in Eq. (23), we compute

$$\hat{\mathbf{A}} = \arg\min_{\mathbf{X}}[-\log L(\mathbf{X})]. \tag{26}$$

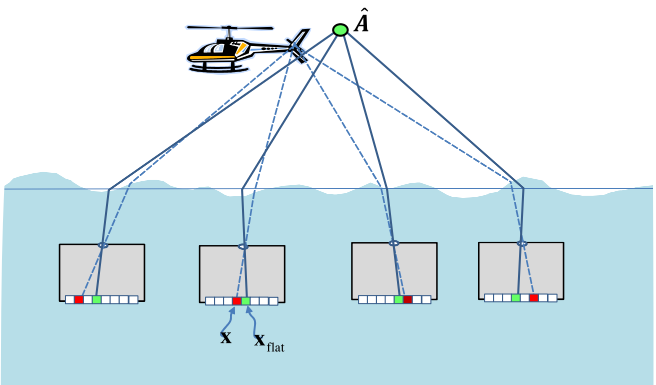

Fig. 9. Airborne position estimation using multiple views. Red pixels are distorted projections of an object point. Green pixels are close to the distorted pixels, while their flat back-projections intersect in 3D at $\hat{\mathbf{A}}$.

Then,

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{X}} \left[ -\sum_{t=1}^{N_{\text{frames}}} \log l^{\mathrm{L}}_t(\mathbf{X}) + \log l^{\mathrm{R}}_t(\mathbf{X}) \right]. \tag{27}$$

The probabilities in Eq. (25) are approximated by Eq. (13). Here $\mathbf{x}^{\mathrm{L}}_{\text{flat}}(\mathbf{X}), \mathbf{x}^{\mathrm{R}}_{\text{flat}}(\mathbf{X})$ are flat pixel projections of a 3D point $\mathbf{X}$ in the right and left views, respectively. These flat projections are *pre-computed once as look-up-tables* for all 3D positions in a grid. For each tracked location $\mathbf{x}^{\mathrm{L}}(t)$ and $\mathbf{x}^{\mathrm{R}}(t)$, we can compute the probabilities $p[\mathbf{x}^{\mathrm{L}}_{\text{flat}}(\mathbf{X})|\mathbf{x}^{\mathrm{L}}(t)]$ and $p[\mathbf{x}^{\mathrm{R}}_{\text{flat}}(\mathbf{X})|\mathbf{x}^{\mathrm{R}}(t)]$ of the corresponding flat pixel $\mathbf{x}^{\mathrm{L}}_{\text{flat}}$ and $\mathbf{x}^{\mathrm{R}}_{\text{flat}}$ for all pixels in the image based on Eqs. (13) and (15). To pass from 3D likelihoods to probabilities in 2D images, Eqs. (15) and (20) yield

$$\hat{\mathbf{A}} = \arg\min_{\mathbf{X}} \left\{ -\sum_{t=1}^{N_{\text{frames}}} \log p[\mathbf{x}^{\mathrm{L}}_{\text{flat}}(\mathbf{X})|\mathbf{x}^{\mathrm{L}}(t)] \right. \\ \left. + \log p[\mathbf{x}^{\mathrm{R}}_{\text{flat}}(\mathbf{X})|\mathbf{x}^{\mathrm{R}}(t)] \right\}. \tag{28}$$

From Eqs. (13) and (15),

$$\log p[\mathbf{x}^{\mathrm{L}}_{\text{flat}}(\mathbf{X})|\mathbf{x}^{\mathrm{L}}(t)] = \log G - \frac{1}{2}[\mathbf{x}^{\mathrm{L}}_{\text{flat}}(\mathbf{X}) \\ - \mathbf{x}^{\mathrm{L}}(t)]^T \left(\Sigma^{\mathrm{L}}_{\mathbf{x}}\right)^{-1}[\mathbf{x}^{\mathrm{L}}_{\text{flat}}(\mathbf{X}) - \mathbf{x}^{\mathrm{L}}(t)]. \tag{29}$$

A similar expression is derived for the right view. Substituting Eq. (29) in Eq. (28),

$$\hat{\mathbf{A}} = \arg\min_{\mathbf{X}} \sum_{t=1}^{N_{\text{frames}}} \\ [\mathbf{x}^{\mathrm{L}}_{\text{flat}}(\mathbf{X}) - \mathbf{x}^{\mathrm{L}}(t)]^T \left(\Sigma^{\mathrm{L}}_{\mathbf{x}}\right)^{-1}[\mathbf{x}^{\mathrm{L}}_{\text{flat}}(\mathbf{X}) - \mathbf{x}^{\mathrm{L}}(t)] \\ + [\mathbf{x}^{\mathrm{R}}_{\text{flat}}(\mathbf{X}) - \mathbf{x}^{\mathrm{R}}(t)]^T \left(\Sigma^{\mathrm{R}}_{\mathbf{x}}\right)^{-1}[\mathbf{x}^{\mathrm{R}}_{\text{flat}}(\mathbf{X}) - \mathbf{x}^{\mathrm{R}}(t)]. \tag{30}$$

For $N_{\text{views}}$ views, Eq. (30) generalizes to

$$\hat{\mathbf{A}} = \arg\min_{\mathbf{X}} \ \mathcal{S}(\mathbf{X}), \tag{31}$$

where

$$\mathcal{S}(\mathbf{X}) = \sum_{t=1}^{N_{\text{frames}}} \sum_{i=1}^{N_{\text{views}}} \\ [\mathbf{x}^{(i)}_{\text{flat}}(\mathbf{X}) - \mathbf{x}^{(i)}(t)]^T \left(\Sigma^{(i)}_{\mathbf{x}}\right)^{-1}[\mathbf{x}^{(i)}_{\text{flat}}(\mathbf{X}) - \mathbf{x}^{(i)}(t)]. \tag{32}$$

We summarize the estimator based on Eq. (30) in Algorithm 2.

**Algorithm 2.** Estimation from Probabilities in Images

```
System Specification: Pre-compute flat pixel projections
    $\mathbf{x}^{\mathrm{L}}_{\mathrm{flat}}(\mathbf{X}), \mathbf{x}^{\mathrm{R}}_{\mathrm{flat}}(\mathbf{X})$ for all 3D positions $\mathbf{X}$ in a grid.
Data: $N_{\mathrm{frames}}$ distorted stereo images of a scene.
Result: Estimated position $\hat{\mathbf{A}}$ of an object at $\mathbf{A}$.
$S = 0$.
for $t = 1$ to$N_{\mathrm{frames}}$ do
    if $t = 1$ then
        Match $\mathbf{x}^{\mathrm{L}}(1)$ and $\mathbf{x}^{\mathrm{R}}(1)$ between views.
    else
        Track the object from time $t-1$ to time $t$ to get $\mathbf{x}^{\mathrm{L}}(t)$
        and $\mathbf{x}^{\mathrm{R}}(t)$.
    end
    $\mathcal{S}(\mathbf{X}) \leftarrow \mathcal{S}(\mathbf{X}) +$
      $[\mathbf{x}^{\mathrm{L}}_{\mathrm{flat}}(\mathbf{X}) - \mathbf{x}^{\mathrm{L}}(t)]^T \left(\Sigma^{\mathrm{L}}_{\mathbf{x}}\right)^{-1} [\mathbf{x}^{\mathrm{L}}_{\mathrm{flat}}(\mathbf{X}) - \mathbf{x}^{\mathrm{L}}(t)] +$
      $[\mathbf{x}^{\mathrm{R}}_{\mathrm{flat}}(\mathbf{X}) - \mathbf{x}^{\mathrm{R}}(t)]^T \left(\Sigma^{\mathrm{R}}_{\mathbf{x}}\right)^{-1} [\mathbf{x}^{\mathrm{R}}_{\mathrm{flat}}(\mathbf{X}) - \mathbf{x}^{\mathrm{R}}(t)]$.
end
$\hat{\mathbf{A}} = \arg\min_{\mathbf{X}} \ \mathcal{S}(\mathbf{X})$.
return $\hat{\mathbf{A}}$.
```

As the location, the experimental uncertainty can also be assessed using probabilities in the 2D images. Using Eqs. (13) and (32), the set $\Psi$ defined in Eqs. (24) is equivalent to

$$\Psi = \{\mathbf{X} : \ [\mathcal{S}(\mathbf{X}) - \mathcal{S}(\hat{\mathbf{A}})] \ < \ \log \tau^{-2}\}, \tag{33}$$

from which the effective spatial support can be computed.

The expressions derived so far are both general and not difficult to implement. Nevertheless, *to gain intuition*, we now describe a restrictive situation. Suppose the $\hat{\mathbf{x}}$ axis of all cameras is horizontal [28] in the lab coordinates. The cameras' $\hat{\mathbf{y}}$ axis can have an arbitrary elevation angle. Then, $\hat{\mathbf{x}} \cdot \hat{\mathbf{Z}} = 0$. If the waves have no dominant direction in the scale of the system, the matrices $\Sigma^{(i)}_{\mathbf{x}}$ are approximately diagonal [28]. The STDs of the projection in the respective image axes are $\sigma_x$ and $\sigma_y$. For identical camcorders with a horizontal baseline looking straight up (not tilted), $\sigma_x = \sigma_y = \sigma$ for all views. Then,

$$\Sigma^{(i)}_{\mathbf{x}} = \sigma^2 \mathbf{I}, \quad \forall i. \tag{34}$$

Substituting Eq. (34) into Eq. (31) leads to

$$\hat{\mathbf{A}} = \arg\min_{\mathbf{X}} \sum_{t=1}^{N_{\mathrm{frames}}} \sum_{i=1}^{N_{\mathrm{views}}} ||\mathbf{x}^{(i)}_{\mathrm{flat}}(\mathbf{X}) - \mathbf{x}^{(i)}(t)||^2. \tag{35}$$

In Eq. (35) the optimal 3D position $\mathbf{X}$ minimizes the distance between all the measured (tracked) distorted image projections $\mathbf{x}^{(i)}(t)$ over all frames. The flat projection associated with $\mathbf{X}$ is $\mathbf{x}^{(i)}_{\mathrm{flat}}(\mathbf{X})$. In a single viewpoint, Eq. (35) yields the 3D location which projects through a flat water to the *center of mass* of all distorted projections.

## 5 Scaling of Uncertainties

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 $\triangle X$ is

$$\triangle X \propto r \triangle x \, , \tag{36}$$

where $\triangle x$ is the pixel size and $r$ is the distance. Viewing monocularly through random refraction distortions, effectively increases the lateral pixel uncertainty $\triangle x$, as expressed in Eq. (13). This increase in $\triangle x$ leads to increase of $\triangle X$ by Eq. (36).

In stereo, there is uncertainty in range 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 $\triangle x$. The range uncertainty [38] is

$$\triangle r \propto \frac{r^2}{b} \triangle x \, . \tag{37}$$

Here $b$ is the effective baseline projected perpendicularly to the optical axis. Again, a wavy WAI, effectively increases $\triangle x$, and accordingly Eq. (37) increases.

The ratio between lateral and depth uncertainties (Eqs. 36 and 37) is

$$\frac{\triangle X}{\triangle r} = \frac{b}{r} \, . \tag{38}$$

Since usually $b \ll r$, it follows that typically $\triangle X \ll \triangle r$. Thus, the uncertainty domain is elongated in the direction away from the camera rig. Relation (38) does not depend on refraction, water waves or our algorithm. It is a general relation that holds for any typical stereo setting, regardless of our environment and method.

*Uncertainty, Number of Views, Number of Time Frames*

As explained in Section 3, multiple uncorrelated measurements reduce the uncertainty of the estimated position. Equations (22) and (31) show that additional views and temporal frames improve the estimation and can be traded-off. Theoretically [39], regardless of refractive distortions, triangulation estimation uncertainty decreases as $1/\sqrt{N_{\mathrm{views}}}$, compared to results obtained using a stereo pair. In general, ML estimation uncertainty decreases as $\approx [N_{\mathrm{views}} N_{\mathrm{frames}}]^{-1/2}$, if the measurements are indeed uncorrelated. As discussed in Section 3.3 and observed in preliminary experiments, views are effectively uncorrelated for a baseline of $\approx 30$ cm, due to short ripples in the WAI. Using the setup described in Section 7, we tracked 40 points in a stereoscopic video. 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. The numbers here are extracted from a specific experiment, and the ratios may vary as a function of setup scale, wave speed and scale. However, ripples caused by short capillary waves move fast and thus decorrelate faster than large waves.

We examined the temporal autocorrelation of each trajectory. Autocorrelation decays in time. The effective correlation time in our experiments was $\approx 9$ frames, equivalent to $\approx 9/30 = 0.3$ seconds. Between consecutive frames, the

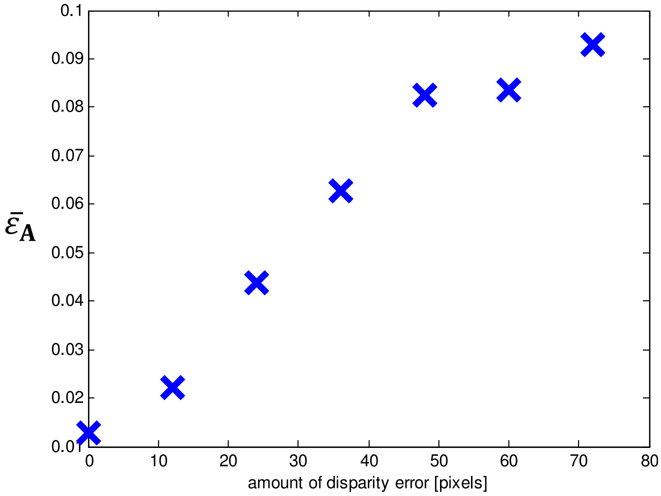

Fig. 10. Error of disparity and its effect on position estimation. The horizontal axis represents disparity error, in pixels. The vertical axis is $\bar{\varepsilon}_{\mathbf{a}}$, relying on $N_{\text{frames}} = 16$, $N_{\text{trials}} = 11$ and 11 different points.

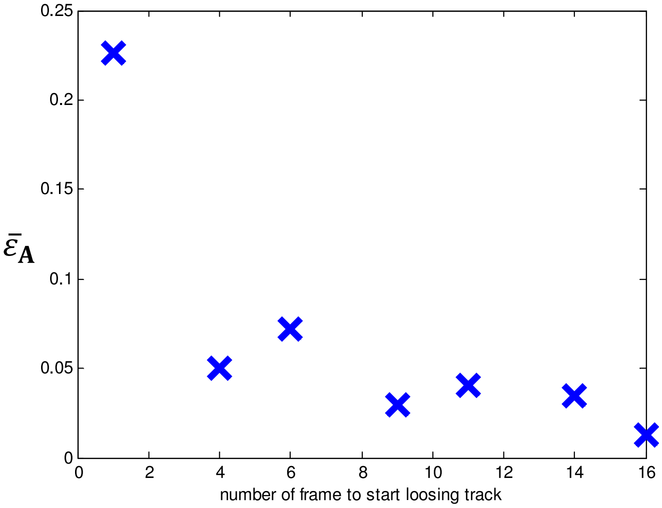

Fig. 11. Undetected tracking failure and its effect on localization error. The horizontal axis represents the frame at which tracking loss occurred in one of the views. The vertical axis is $\bar{\varepsilon}_{\mathbf{a}}$, using $N_{\text{frames}} = 16$.

correlation coefficient is $\approx 0.6$. Hence, solely using time to decorrelate measurements, video rate is inefficiently too fast for wave motion. This 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 (as described), faster than sequential monocular imaging.

## 6 Simulations

We performed simulations based on our experimental system parameters. Similarly to [26], the water surface was simulated using a sea surface model from [40] for each time instance. We set the wind to be 2.5 m/s and a peak-to-peak amplitude of 2 cm. A synthetic stereo camera rig of a 27.5 cm baseline was set to be at a depth of 15 cm. A planar object was placed 165 cm above the water surface ($A_z = 180$ cm). By ray tracing, we synthesize water distorted image pairs, changing the water surface in each temporal frame. These frame pairs are the input to our algorithm. Simulations were repeated for $N_{\text{trials}}$ distinct random wave conditions.

Denote the average relative estimation error by

$$\bar{\varepsilon}_{\mathbf{A}} = \frac{1}{N_{\text{trials}}\|\mathbf{A}\|_2} \sum_{m=1}^{N_{\text{trials}}} \|\mathbf{A} - \hat{\mathbf{A}}(m)\|_2 \,. \tag{39}$$

Specifically, for $N_{\text{trials}} = 30$ and $N_{\text{frames}} = 1$, $\bar{\varepsilon}_{\mathbf{A}} = 0.34$ m. This is far larger than the 2 cm length of the cubic voxels we used. Hence, in the most basic estimation, this voxelization is not the dominant source of uncertainty. Next, we plot simulations quantified by $\bar{\varepsilon}_{\mathbf{A}}$.

Correspondence error yields erroneous disparity. The consequence of horizontal pixel disparity error $\triangle x$ between L and R views is plotted in Fig. 10. As expected, a high disparity error increases the triangulation estimation error.

Imperfect tracking is a source of error, in case video is used. Fig. 11 shows the effect of major tracking failure on position estimation. We tested situations where tracking is lost, and the tracker is unaware of the loss. Then, location estimation proceeds using wrongly tracked points. If the tracking is lost at the beginning, $\bar{\varepsilon}_{\mathbf{A}}$ is large. If tracking is lost after a larger number of frames, the relative error is much smaller. Now, suppose the tracker detects loss and thus aborts. In such a system, the estimation uses only $N_{\text{frames}}$ frames where tracking was fine, per object point. Fig. 12 plots $\bar{\varepsilon}_{\mathbf{A}}$ as a function of $N_{\text{frames}}$, for $N_{\text{trials}} = 11$ simulations. A $1/\sqrt{N_{\text{frames}}}$ falloff curve is included. The simulations generally follow this trend, particularly when $N_{\text{frames}}$ is small. However, the simulated falloff starts to flatten at large $N_{\text{frames}}$. The reason is that the fixed cubic voxel size of 2 cm length has increasing significance. As expected, we found in simulations that there is no benefit in using very small voxels, when the likelihood spread caused by WAI waves is very large. Voxels can be large, as long as they are smaller than the effective spread of $L$. As $N_{\text{frames}}$ increases, fixed voxelization eventually becomes a resolution bottleneck. This may be improved by more enhanced algorithms in which the voxel size adapts as $N_{\text{frames}}$ increases.

Calibration error is simulated by gradually changing the position of the right camera from its correct position.

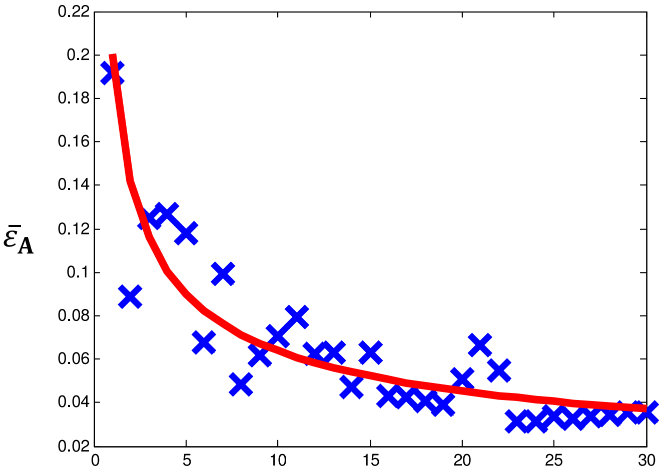

Fig. 12. Tacking is lost and aborted. The plot shows the effect on location estimation. The horizontal axis represents the number of frames until tracking is lost, used for estimation. The vertical axis is $\bar{\varepsilon}_{\mathbf{a}}$ based on $N_{\text{trials}} = 11$ and 11 different points.

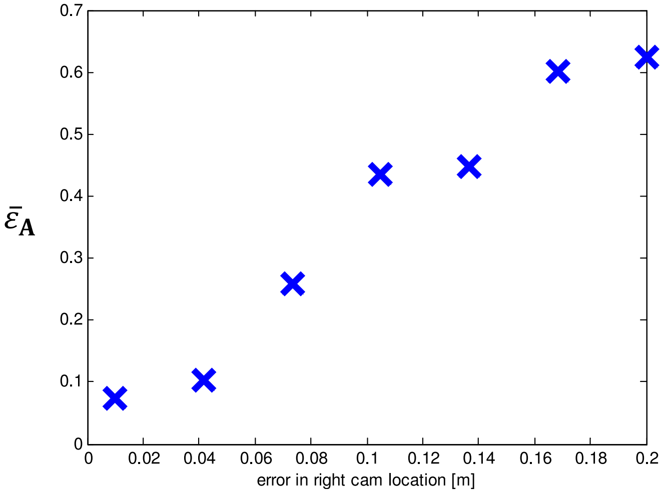

Fig. 13. Error in calibration of the stereo baseline (horizontal plot axis), and its effect on estimation. The vertical axis is $\bar{\varepsilon}_{\mathbf{a}}$, based on $N_{\text{frames}} = 16$.

Fig. 13 plots the increase of the estimation error with the calibration error.

## 7 Experiments

We conducted experiments in a 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. 14 shows the lab setup. The video sequences of the two cameras were synchronized in post processing, by detecting light strobes that we had recorded [41], [25]. 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 [11], [17], [42]. The Matlab calibration toolbox [43] both calibrated this system and verified measurements at other underwater distances, in [25]. 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 $h_{\text{pixel}}$ supplied by the camera manufacturer.

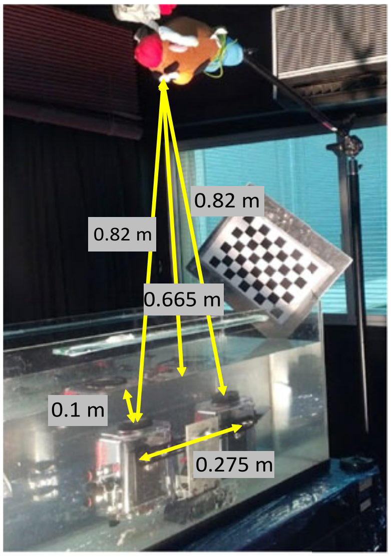

Fig. 14. Lab experiment setup. Camcorders in a stereo rig are submerged in a water tank. Objects of interest are hanged in the air above, at known locations that serve as ground truth.

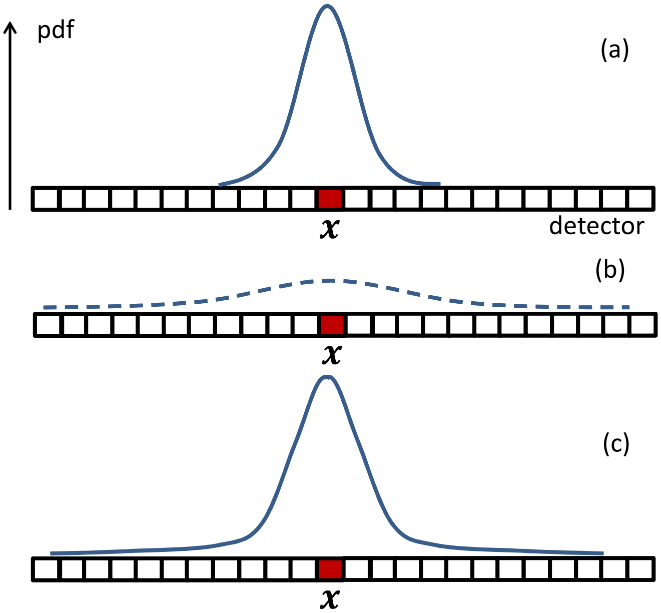

Fig. 15. Pixel position statistical model. [a] Normal distribution of the distorted pixel location, around its undistorted location. [b] A long tail distribution, created by a Gaussian having a very wide support. [c] Weighted average of the distributions in [a] and [b].

We used Large Displacement Optical Flow [33], [34] for tracking interest points.[3] We found this tracker to be rather robust to the harsh and fast distortions exhibited. Outliers in $\mathbf{x}$ may cause numerical problems in 3D likelihood estimation. To make Algorithm 1 robust to outliers, we incorporated a long tail into the modeled Gaussian, as illustrated in Fig. 15. This was achieved by setting the distortion PDF to be

$$p(\mathbf{d}) = \alpha \mathcal{N}(0, \Sigma_{\mathbf{x}}) + (1 - \alpha)\mathcal{N}(0, \mathbf{P}), \tag{40}$$

where $\mathbf{P}$ is a diagonal matrix, expressing a Gaussian whose axial STDs are respectively $7\sigma_x$ and $7\sigma_y$. We used $\alpha = 0.98$. In all experiments, we used fixed, uniform cubic voxels, 2 cm long. A Gaussian was fitted to the tracked projections, over several frames, to empirically estimate $\Sigma_{\mathbf{x}}$. In Eqs. (24) and (33), the value $\tau = 0.01$ was used.

### 7.1 Triangulation in a Laboratory Experiment

In the lab, the camera pair was placed in a water tank at a depth of $Z_{\text{flat}} = 10$ cm, looking upwards through the WAI (Fig. 14). Water waves were created by hand, producing random uncontrolled waves. The scene includes two objects. The potato head doll is $\approx 82$ cm from the cameras and the swan photo is $\approx 133$ cm from the cameras. First, we deal with the potato head. Sample pair frames are shown in Figs. 16a, 16b. The 3D domain was first set to a volume of $0.8 \times 0.8 \times 4\,\text{m}^3$, which projects to an image area corresponding to $\approx 260 \times 200$ pixels. Using $N_{\text{frames}} = 3$ and Algorithm 1, the 3D object position was estimated, as shown in Fig. 16c. Algorithm 2 proved to be significantly faster, allowing us to use a larger volume of $3 \times 3 \times 9\,\text{m}^3$, which projects to the full

[3] Recently, also a Locally Orderless Tracking (LOT) [44] demonstrated successful tracking under refractive distortions.

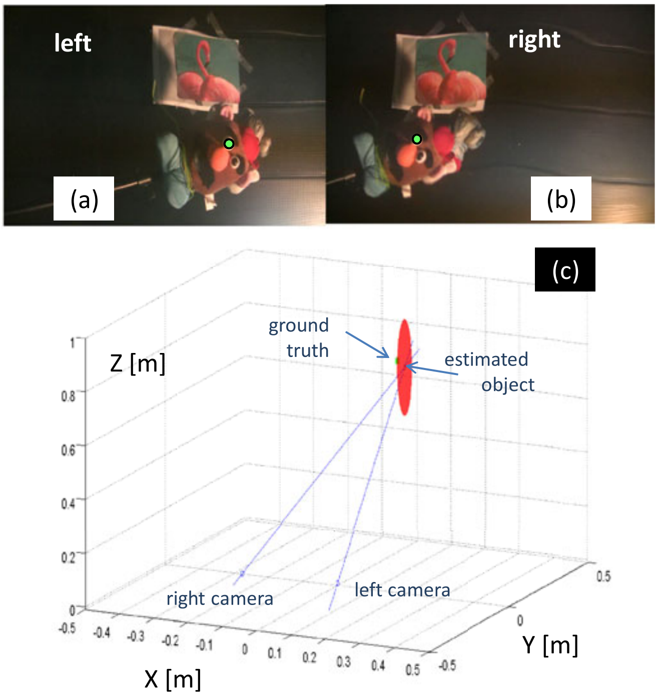

Fig. 16. Results of a lab experiment. [a] Sample frame from the left viewpoint. [b] Sample frame from the right viewpoint. [c] The triangulation results in lab coordinates. The ellipsoid represents the uncertainty of the result.

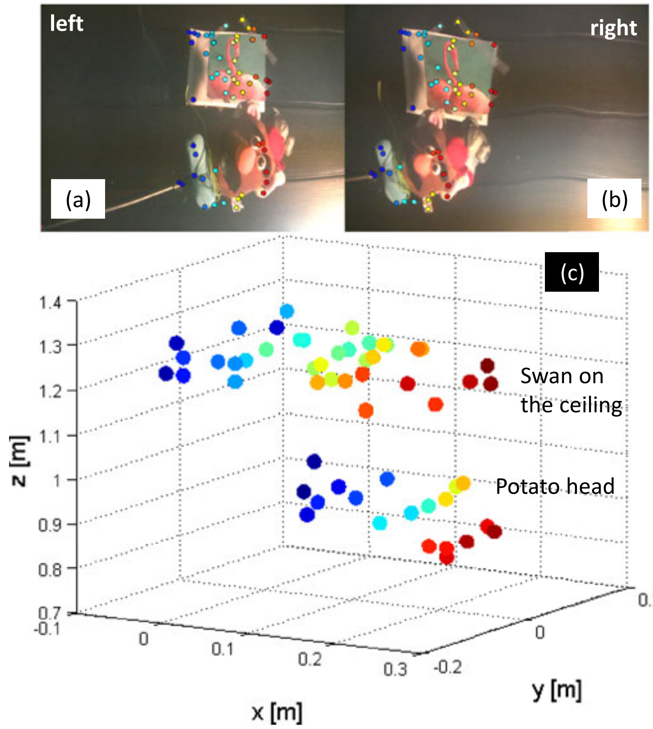

Fig. 17. Results of a lab experiment. [a] Sample frame from the left viewpoint with automatically computed matches overlayed. [b] Sample frame from the right viewpoint with automatically computed matches overlaid. [c] The triangulation results in lab coordinates for multiple points. The colors match the points in [a] and [b].

frame of $540 \times 720$ pixels. As shown in Table 1, the results are consistent with the ground-truth, up to the experimental uncertainty.

To triangulate multiple points, matches were automatically found between the left and the right views using [19], as described in Section 2.3. The obtained matching (disregarding the black ceiling) is shown in Figs. 17a and 17b. The triangulation results are shown in Fig. 17c. The points on the potato head doll are below the points on the swan. This is expected since the swan is attached to the ceiling. The STD of $Z$ across all points on the swan is $3.9$ cm. Refer to Table 1 for all results.

## 7.2 Pool Experiment

We performed a similar experiment at an indoor swimming pool. The stereoscopic camera rig was mounted on a tripod and submerged. Several objects resided above the cameras

TABLE 1
Summary of Experimental Results [cm]

| | Lab Potato | Lab Swan | Pool Potato | Pool Red Broom | Pool Blue Broom |
|---|---|---|---|---|---|
| Ground Truth<br>$\begin{pmatrix} A_x \\ A_y \\ A_z \end{pmatrix}$ | $\begin{pmatrix} 14 \\ 8 \\ 82 \end{pmatrix} \pm 1$ | $A_z = 133 \pm 1$ | $A_z = 166 \pm 1$ | $A_z = 133 \pm 1$ | $A_z = 174 \pm 1$ |
| Algorithm 1<br>$\begin{pmatrix} A_x^{\min}, & \hat{A}_x, & A_x^{\max} \\ A_y^{\min}, & \hat{A}_y, & A_y^{\max} \\ A_z^{\min}, & \hat{A}_z, & A_z^{\max} \end{pmatrix}$ | $\begin{pmatrix} 13, & \mathbf{14}, & 15 \\ 7, & \mathbf{8}, & 9 \\ 80, & \mathbf{82}, & 84 \end{pmatrix}$ | $\begin{pmatrix} 13, & \mathbf{14}, & 15 \\ -9, & \mathbf{-8}, & -7 \\ 122, & \mathbf{128}, & 138 \end{pmatrix}$ | $\begin{pmatrix} 17, & \mathbf{18}, & 20 \\ 8, & \mathbf{10}, & 14 \\ 144, & \mathbf{160}, & 190 \end{pmatrix}$ | $\begin{pmatrix} 39, & \mathbf{40}, & 41 \\ -7, & \mathbf{-6}, & -4 \\ 120, & \mathbf{128}, & 130 \end{pmatrix}$ | $\begin{pmatrix} 5, & \mathbf{6}, & 7 \\ 3, & \mathbf{4}, & 5 \\ 168, & \mathbf{178}, & 186 \end{pmatrix}$ |
| Algorithm 2<br>$\begin{pmatrix} A_x^{\min}, & \hat{A}_x, & A_x^{\max} \\ A_y^{\min}, & \hat{A}_y, & A_y^{\max} \\ A_z^{\min}, & \hat{A}_z, & A_z^{\max} \end{pmatrix}$ | $\begin{pmatrix} 13, & \mathbf{14}, & 15 \\ 7, & \mathbf{8}, & 9 \\ 76, & \mathbf{78}, & 82, \end{pmatrix}$ | $\begin{pmatrix} 13, & \mathbf{14}, & 15 \\ -10, & \mathbf{-8}, & -6 \\ 118, & \mathbf{128}, & 140 \end{pmatrix}$ | $\begin{pmatrix} 17, & \mathbf{18}, & 20 \\ 8, & \mathbf{10}, & 12 \\ 140, & \mathbf{154}, & 174 \end{pmatrix}$ | $\begin{pmatrix} 42, & \mathbf{44}, & 46 \\ -7, & \mathbf{-6}, & -4 \\ 116, & \mathbf{134}, & 144 \end{pmatrix}$ | $\begin{pmatrix} 4, & \mathbf{6}, & 7 \\ 2, & \mathbf{4}, & 6 \\ 154, & \mathbf{172}, & 204 \end{pmatrix}$ |

*The table presents the results including ground truth and uncertainty. Boldface numbers represent the estimated location $\widehat{A}$ while the uncertainty bounds are shown on both sides*

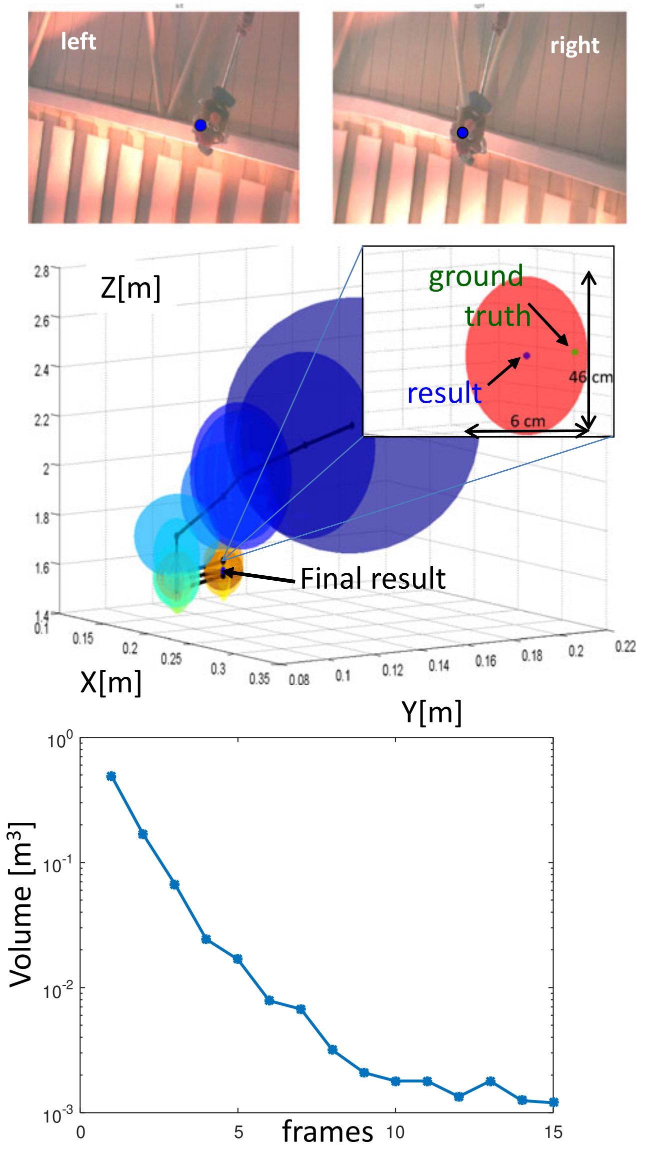

Fig. 18. Results of a pool experiment, triangulating a potato head. [Top] Sample frames from the left and the right viewpoints. [Middle] The illustration shows the estimation evolving as $t$ and $N_{\text{frames}}$ increase. Ellipsoids represent the uncertainty of the result. Time is color coded: blue to red. A zoom-in shows the final result and the ground truth. [Bottom] The uncertainty volume plotted as a function of time.

at $A_z \in [1, 2]$m. One object (potato head doll) was placed 1.6m above the rig. Algorithm 1 yielded results evolving over time $t$, as $N_{\text{frames}}$ in Eq. (21) increased from 1 to 15.

The uncertainty of the estimation gradually decreased. This is illustrated in Fig. 18: the color coding of the ellipsoids indicates temporal evolution of the estimation: blue to red. We applied Algorithm 2. The results are shown in Fig. 19. The estimated location of the eye of the potato head doll is marked at the bottom.

In another scenario, two brooms were placed above the cameras at two different positions. Sample frames and the estimated positions are shown in Fig. 20 and in Table 1. The result was obtained after 16 frames. Each new frame introduces additional information. Thus, the support of the 3D likelihood function shrinks over time. The estimated results (see Table 1) are consistent with the ground truth.

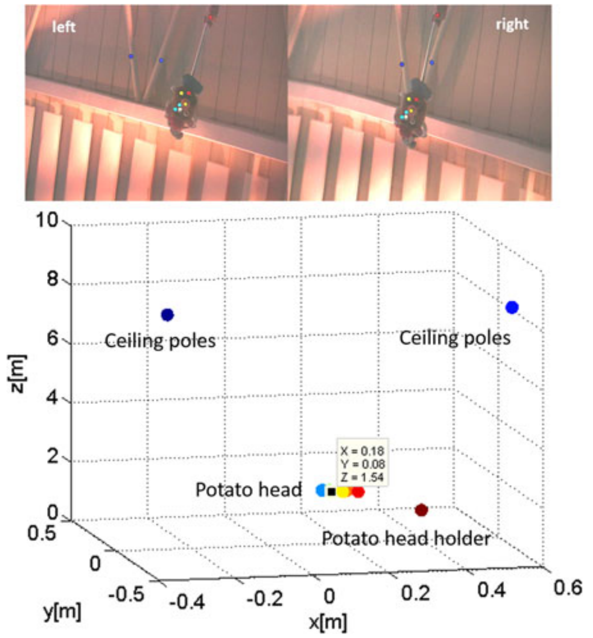

Fig. 19. Results of a pool experiment, triangulating a potato head for several points on the doll and on the ceiling. [Top] Left and right views with corresponding, automatically computed, points overlaid. [Bottom] Triangulation result for the corresponding points. The estimated location of the eye of the doll (green point) is shown.

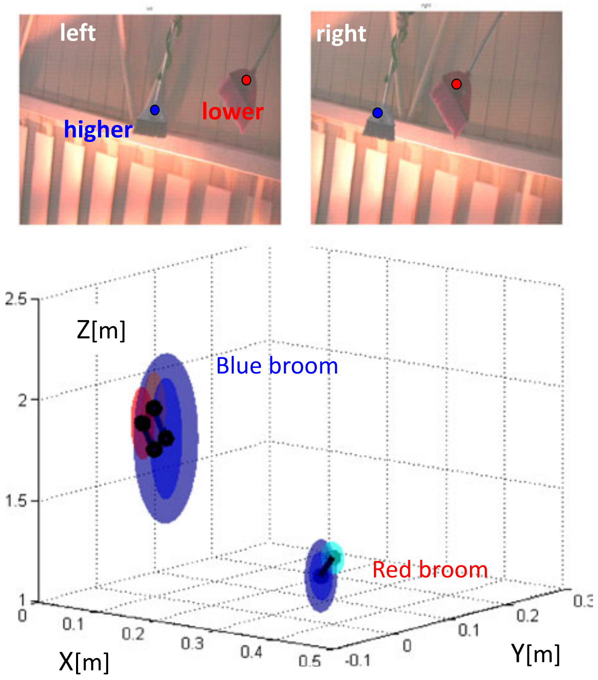

Fig. 20. A pool experiment, triangulating brooms. The result is given in lab coordinates. Ellipsoids represent the uncertainty of the result. Time is represented by color coding of the ellipsoids: blue to red.

## 8 Discussion

This work deals with triangulation through highly random refractive distortions. The formulation is stochastic. After the definition of the problem, we present two algorithmic solutions. In addition, automatic image matching obtains multiple correspondences across views. The automated correspondence provides initialization for efficient triangulation. Although the 3D ML algorithm is intuitive, the 2D algorithm which is based directly on image plane coordinates, simplifies the computations and reduces run time. Thus, it allows triangulation for multiple points. We use the method from [19] to obtain multiple point correspondences between views. The problem and principles presented here are not limited to the specific algorithms we used to demonstrate triangulation. Improvements may stem from adaptive voxelization, other parametric windowing, better trackers, advanced and future matching algorithms designed for severe distortions.

The approach may generalize to detection and analysis of moving objects, through dynamic random refraction. This task was treated in a monocular view [28], but has yet to exploit multiview data. It may be possible to generalize the system to light-field or integral imaging [45], [46], [47], and thus acquire many images simultaneously.

There are monocular methods to "flatten" images taken through a wavy WAI, which do not triangulate objects and measure their range. Lucky imaging [48], [49], [50], [51] requires a large number of frames, out of which a best representative frame is selected per patch. Less temporal frames are needed if a spatiotemporal distortion model [52], [53], [54] is fitted to the image data, or given an undistorted template image [55]. A virtual periscope is theoretically proposed [56] based on a wave model constrained by self occlusions. Possibly, rectification can be based on WAI estimates derived from dedicated optical measurements [57], [58], [59], [60], [61], including the STELLA MARIS virtual periscope approach [62].

Such methods can be useful: they can estimate rectified images in a first step. The rectified images can be triangulated in a second step. On the other hand, our stochastic triangulation, which directly handles raw multiview data has advantages. First, random errors in this difficult rectification task generally challenge deterministic triangulation, since erroneous LOSs may not intersect in 3D. Second, a flattened monocular image may harbor a global bias: a very smooth but slanted WAI yields correct monocular content having a spatial offset that biases triangulation.

We demonstrated our triangulation approach using a submerged system. However, the approach can also be applied in the opposite case, of an airborne camera looking into water. There are multi-camera methods dedicated to recovering a wavy WAI [23], [63], [64], [65], [66]. These methods, however, often rely on a known calibration target being the observed object. They were thus not intended to triangulate objects behind a WAI.

The method can possibly be used through atmospheric turbulence [67], [68], [69]. It is worth noting that parallel image acquisition was suggested by [70] for deconvolution-favorable imaging while [71], [72] used multi view imaging to recover the 3D structure of atmospheric turbulence.

## ACKNOWLEDGMENTS

We thank the anonymous reviewers for their useful comments and Joseph Shamir for useful discussions. We thank Daniel Yagodin for his help in the lab experiment, and the swimming pool facilities of the Technion for letting us conduct the experiments. We thank the authors of [19] for letting us use their image matching code. Yoav Schechner is a Landau Fellow - supported by the Taub Foundation. This research was mainly conducted while Marina Alterman was with the Department of Electrical Engineering, Technion - Israel Inst. Technology. 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] G. Katzir and N. Intrator, "Striking of underwater prey by a reef heron, egretta gularis schistacea," *J. Compar. Phys. A.*, vol. 160, pp. 517–523, 1987.

[2] A. Ben-Simon, O. Ben-Shahar, G. Vasserman, M. Ben-Tov, and R. Segev, "Visual acuity in the archerfish: Behavior, anatomy, and neurophysiology," *J. Vis.*, vol. 12, no. 12, 2012.

[3] Y. Y. Schechner, "A view through the waves," *Marine Tech. Society J.*, vol. 47, pp. 148–150, 2013.

[4] Y. Adato, Y. Vasilyev, T. Zickler, and O. Ben-Shahar, "Shape from specular flow," *Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 2054–2070, Nov. 2010.

[5] K. Ikeuchi, M. Sakauchi, H. Kawasaki, and I. Sato, "Constructing virtual cities by using panoramic images," *Int. J. Comput. Vision*, vol. 58, no. 3, pp. 237–247, 2004.

[6] R. I. Hartley and P. Sturm, "Triangulation," *Comput. Vis. Image Understanding*, vol. 68, no. 2, pp. 146–157, 1997.

[7] S. Peleg, M. Ben-Ezra, and Y. Pritch, "Omnistereo: Panoramic stereo imaging," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2001, vol. 23, no. 3, pp. 279–290.

[8] N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: Exploring image collections in 3D," in *Proc. SIGGRAPH*, 2006, pp. 835-846.

[9] Y. Tsin, S. Kang, and R. Szeliski, "Stereo matching with reflections and translucency," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2003, vol. 1, pp. I–702.

[10] Y. J. Chang and T. Chen, "Multi-view 3D reconstruction for scenes under the refractive plane with known vertical direction," in *Proc. Int. Conf. Comput. Vision*, 2011, pp. 351–358.

[11] V. Chari and P. Sturm, "Multi-view geometry of the refractive plane," in *Proc. BMVC*, 2009, pp. 56.1–56.11.

[12] R. Ferreira, J. Costeira, and J. Santos, "Stereo reconstruction of a submerged scene," in *Proc. Pattern Recog. Image Anal.*, 2005, vol. 3522, pp. 102–109.

[13] M. Gupta, S. Narasimhan, and Y. Schechner, "On controlling light transport in poor visibility environments," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2008, pp. 1–8.

[14] G. Horváth and D. Varjú, "On the structure of the aerial visual field of aquatic animals distorted by refraction," *Bull. Math. Bio.*, vol. 53, no. 3, pp. 425–441, 1991.

[15] S. Koppal, I. Gkioulekas, T. Zickler, and G. Barrows, "Wide-angle micro sensors for vision on a tight budget," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2011, pp. 361–368.

[16] H. Saito, H. Kawamura, and M. Nakajima, "3D shape measurement of underwater objects using motion stereo," in *Proc. IEEE IECON*, 2002, vol. 2, pp. 1231–1235.

[17] T. Treibitz, Y. Y. Schechner, and H. Singh, "Flat refractive geometry," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2008, pp. 1–8.

[18] F. Dellaert, S. Seitz, C. Thorpe, and S. Thrun, "Structure from motion without correspondence," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2000, vol. 2, pp. 557–564.

[19] Y. Lipman, S. Yagev, R. Poranne, D. Jacobs, and R. Basri, "Feature matching with bounded distortion," *ACM Trans. Graph.*, vol. 33, no. 3, p. 26, 2014.

[20] M. Alterman, Y. Schechner, and Y. Swirski, "Triangulation in random refractive distortions," in *Proc. IEEE Int. Conf. Comput. Photography*, 2013, pp. 1–10.

[21] I. Ihrke, K. Kutulakos, H. P. Lensch, M. A. Magnor, and W. Heidrich, "State of the art in transparent and specular object reconstruction," in *Proc. Eurographics*, 2008, pp. 87–108.

[22] 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*, vol. 19, no. 4, pp. 687–694, 2002.

[23] N. Morris and K. Kutulakos, "Dynamic refraction stereo," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2005, vol. 2, pp. 1573–1580.

[24] Y. Y. Schechner and N. Karpel, "Attenuating natural flicker patterns," in *Proc. MTS/IEEE Oceans*, 2004, pp. 1262–1268.

[25] Y. Swirski and Y. Y. Schechner, "3Deflicker from motion," in *Proc. IEEE ICCP*, 2013, pp. 1–9.

[26] Y. Swirski, Y. Y. Schechner, B. Herzberg, and S. Negahdaripour, "CauStereo: Range from light in nature," *App. Opt.*, vol. 50, pp. 89–101, 2011.

[27] Y. Swirski, Y. Y. Schechner, and T. Nir, "Variational stereo in dynamic illumination," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 1124–1131.

[28] M. Alterman, Y. Y. Schechner, J. Shamir, and P. Perona, "Detecting motion through dynamic refraction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 245–251, Jan. 2013.

[29] M. J. Kidger, *Fundamental Optical Design*. Bellingham, WA, USA: SPIE Press, 2002.

[30] H. Kleerekoper and T. Malar, "Orientation through sound in fishes," in *Ciba Foundation Symposium - Hearing Mechanisms in Vertebrates*. Hoboken, NJ, USA: Wiley, 2008, pp. 187–206.

[31] J. B. Steen, "The swim bladder as a hydrostatic organ," *Fish Physiology: The Nervous Syst., Circulation, Respiration*, vol. 4, pp. 413–443, 1970.

[32] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[33] T. Brox. (2011). Large displacement optical flow code [Online]. Available: http://132.230.167.110/people/brox/resources/pami2010Matlab.zip.

[34] T. Brox and J. Malik, "Large displacement optical flow: Descriptor matching in variational motion estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 500–513, Mar. 2011.

[35] C. Cox and W. Munk, "Statistics of the sea surface derived from sun glitter," *J. Mar. Res.*, vol. 13, pp. 198–227, 1954.

[36] M. B. Hullin, H. P. A. Lensch, R. Raskar, H.-P. Seidel, and I. Ihrke, "Dynamic display of BRDFs," in *Proc. Eurographics*, 2011, pp. 475–483.

[37] A. Molkov and L. Dolin, "Determination of wind roughness characteristics based on an underwater image of the sea surface," *Izvestiya, Atmospheric Oceanic Phys.*, vol. 48, pp. 552–564, 2012.

[38] D. Gallup, J.-M. Frahm, P. Mordohai, and M. Pollefeys, "Variable baseline/resolution stereo," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2008, pp. 1–8.

[39] C. Fraser, "Network design considerations for non-topographic photogrammetry," *Photogrammetric Eng. Remote Sensing*, vol. 50, no. 8, pp. 1115–1126, 1984.

[40] J. Miranda, A. Camps, J. Gomez, M. Vall-llossera, and R. Villarino, "Time-dependent sea surface numerical generation for remote sensing applications," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2005, pp. 2527–2530.

[41] D. Bradley, B. Atcheson, I. Ihrke, and W. Heidrich, "Synchronization and rolling shutter compensation for consumer video camera arrays," *in Proc. IEEE Int. Workshop Projector-Camera Syst.*, 2009, pp. 1–8.

[42] A. Jordt-Sedlazeck and R. Koch, "Refractive calibration of underwater cameras," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 846–859.

[43] J. Y. Bouguet, "Camera calibration toolbox for matlab," www.vision.caltech.edu/bouguetj/calib_doc. [Online]. Available: www.vision.caltech.edu/bouguetj/calib_doc

[44] S. Oron, A. Bar-Hillel, D. Levi, and S. Avidan, "Locally orderless tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2012, pp. 1940–1947.

[45] S. H. Hong, J. S. Jang, and B. Javidi, "Three-dimensional volumetric object reconstruction using computational integral imaging," *Optics Express*, vol. 12, no. 3, pp. 483–491, 2004.

[46] Raytrix, [Online]. Available: http://www.raytrix.de.

[47] G. Wetzstein, D. Roodnick, R. Raskar, and W. Heidrich, "Refractive shape from light field distortion," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 1180–1186.

[48] A. Donate and E. Ribeiro, "Improved reconstruction of images distorted by water waves," in *Proc. Adv. Comp. Graph. Comp. Vis.*, 2007, pp. 264–277.

[49] A. Efros, V. Isler, J. Shi, and M. Visontai, "Seeing through water," *Proc. Adv. Neural Inf. Process. Syst.*, 2004, vol. 17, pp. 393–400.

[50] 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 *Proc. IEEE OCEANS*, 2008, pp. 1–12.

[51] Z. Wen, A. Lambert, D. Fraser, and H. Li, "Bispectral analysis and recovery of images distorted by a moving water surface," in *Proc. JOSA A*, 2010, vol. 49, no. 33, pp. 6376–6384.

[52] H. Murase, "Surface shape reconstruction of a nonrigid transport object using refraction and motion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 10, pp. 1045–1052, Oct. 1992.

[53] O. Oreifej, G. Shu, T. Pace, and M. Shah, "A two-stage reconstruction approach for seeing through water," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2011, pp. 1153–1160.

[54] Y. Tian and S. G. Narasimhan, "Seeing through water: Image restoration using model-based tracking," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, pp. 2303–2310.

[55] Y. Tian, "A globally optimal data-driven approach for image distortion estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2010, pp. 1277–1284.

[56] 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 Random Complex Media*, vol. 16, pp. 521–530, 2006.

[57] S. Baglio, C. Faraci, and E. Foti, "Structured light approach for measuring sea ripple characteristics," in *Proc. IEEE OCEANS*, 1998, vol. 1, pp. 449–453.

[58] D. Dabiri and M. Gharib, "Interaction of a shear layer with a free surface," in *Proc. Int. Sympos. Flow Vis.*, 2000, pp. 72:1–72:12.

[59] L. S. Dolin, A. G. Luchinin, V. I. Titov, and D. G. Turlaev, "Correcting images of underwater objects distorted by sea surface roughness," in *Proc. SPIE*, 2007, vol. 6615.

[60] B. Jähne, J. Klinke, and S. Waas, "Imaging of short ocean wind waves: A critical theoretical review," *JOSA A*, vol. 11, pp. 2197–2209, 1994.

[61] H. Schultz and A. Corrada-Emmanuel, "System and method for imaging through an irregular water surface," U.S. Patent 7,630,077, 2007.

[62] M. Alterman, Y. Swirski, and Y. Schechner, "STELLA MARIS: Stellar marine refractive imaging sensor," in *Proc. IEEE ICCP*, 2014, pp. 1–10.

[63] Y. Ding, F. Li, Y. Ji, and J. Yu, "Dynamic 3D fluid surface acquisition using a camera array," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 2478–2485.

[64] M. S. Schmalz, "Integration of stereophotogrammetry with image restoration models for distortion-tolerant viewing through the sea surface," in *Proc. SPIE*, 1993, vol. 1943, pp. 115–128.

[65] H. Schultz, "Shape reconstruction from multiple images of the ocean surface," *Surface, Photogrammetric Eng. Remote Sens.*, vol. 62, pp. 93–99, 1994.

[66] R. Westaway, S. Lane, and D. Hicks, "Remote sensing of clear-water, shallow, gravel-bed rivers using digital photogrammetry," *Photogrammetric Eng. Remote Sens.*, vol. 67, no. 11, pp. 1271–1282, 2001.

[67] N. Joshi and M. F. Cohen, "Seeing Mt. Rainier: Lucky imaging for multi-image denoising, sharpening, and haze removal," in *Proc. IEEE Conf. Compute. Photography*, 2010, pp. 28–30.

[68] Y. Tian, S. Narasimhan, and A. Vannevel, "Depth from optical turbulence," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2012, pp. 246–253.

[69] X. Zhu and P. Milanfar, "Removing atmospheric turbulence via space-invariant deconvolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 157–170, Jan. 2013.

[70] M. Loktev, O. Soloviev, S. Savenko, and G. Vdovin, "Speckle imaging through turbulent atmosphere based on adaptable pupil segmentation," *Opt. Lett.*, vol. 36, no. 14, pp. 2656–2658, Jul. 2011.

[71] M. Alterman, Y. Y. Schechner, M. Vo, and S. G. Narasimhan, "Passive tomography of turbulence strength," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 47–60.

[72] 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 *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 767–782.

**Marina Alterman** received the BSc, MSc, and PhD degrees in electrical engineering from the Technion-Israel Institute of Technology, Haifa, Israel, in 2006, 2010, and 2015 respectively. She is currently a postdoctoral researcher in the Comp. Photo. Lab at Northwestern University. She was the recipient of the Ollendorff Fellow Scholarship in 2014 and the IMVC second best paper award in 2014. Her research involves physics-based computer vision.

**Yoav Y. Schechner** received the BA and MSc degrees in physics and the PhD degree in electrical engineering from the Technion-Israel Institute of Technology in 1990, 1996, and 2000, respectively. During the years 2000 to 2002 he was a research scientist at the Computer Science Department in Columbia University. Since 2002, he has been a faculty member at the Department of Electrical Engineering of the Technion, where he heads the Hybrid Imaging Lab. From 2010 to 2011 he was a visiting Scientist at Caltech and NASA's Jet Propulsion Laboratory. His research is focused on computer vision, the use of optics and physics in imaging and computer vision, and on multi-modal sensing. He was the recipient of the Wolf Foundation Award for Graduate Students in 1994, the Guttwirth Special Distinction Fellowship in 1995, the Ollendorff Award in 1998, the Morin Fellowship in 2000-2002, the Landau Fellowship in 2002-2004 and the Alon Fellowship in 2002-2005. He received the Klein Research Award in 2006, the Outstanding Reviewer Awards in IEEE CVPR 2007 and ICCV 2007 and the Best Paper Award in IEEE ICCP 2013. He is a member of the IEEE.

**Yohay Swirski** received the BSc degree in electrical engineering and physics, and the PhD degree in electrical engineering from the Technion-Israel Institute of Technology, Haifa, Israel, in 2006 and 2012, respectively. He was the recipient of the Ollendorff Fellow Scholarship in 2013. His research involves underwater physics-based computer vision. He is also a PADI Divemaster.