Short Papers

Detecting Motion through Dynamic Refraction

Marina Alterman,
Yoav Y. Schechner, *Member*, *IEEE*,
Pietro Perona, *Member*, *IEEE*, and
Joseph Shamir, *Life Fellow*, *IEEE*

Abstract—Refraction causes random dynamic distortions in atmospheric turbulence and in views across a water interface. The latter scenario is experienced by submerged animals seeking to detect prey or avoid predators, which may be airborne or on land. Man encounters this when surveying a scene by a submarine or divers while wishing to avoid the use of an attention-drawing periscope. The problem of inverting random refracted dynamic distortions is difficult, particularly when some of the objects in the field of view (FOV) are moving. On the other hand, in many cases, just those moving objects are of interest, as they reveal animal, human, or machine activity. Furthermore, detecting and tracking these objects does not necessitate handling the difficult task of complete recovery of the scene. We show that moving objects can be detected very simply, with low false-positive rates, even when the distortions are very strong and dominate the object motion. Moreover, the moving object can be detected even if it has zero mean motion. While the object and distortion motions are random and unknown, they are mutually independent. This is expressed by a simple motion feature which enables discrimination of moving object points versus the background.

Index Terms—Motion detection, refraction, random media, classification, distortion

1 Introduction

Changes in refractive index of a medium along a line of sight (LOS) cause distortion [21], [31], [41]. Such changes are created by atmospheric turbulence. They are also created when looking through a water-air interface (WAI). In a wavy WAI, the refracted view changes spatiotemporally, as in a turbulent atmosphere but more strongly, even in mild weather conditions. In either case, the result is random dynamic distortions. The distortions are particularly strong when a submerged viewer looks obliquely upward through a WAI (Figs. 1 and 2). Due to Snell's Law, small changes in the WAI slope are amplified to large angular changes of the airborne LOS, particularly near the horizon.

The paper deals with vision under natural random distortions. We focus on shallow oblique views upward through a wavy WAI for several reasons. First, it is the most challenging since the distortions in this scenario are strongest. Second, it is related to biological vision: It is experienced by submerged animals seeking to detect prey or avoid predators which are airborne or on land. The archer fish [40] shoots down airborne prey using a water jet, aiming while being submerged. It has also been hypothesized that seals watch for ambushing polar bears, prior to surfacing for air. There are also airborne and land animals [25] that detect and hunt fish by looking down through the WAI. Oblique upward vision through a WAI can function as a virtual periscope for submariners and divers who wish to avoid using attention-drawing physical periscopes.

Computer vision in random distortions has been studied extensively, as we detail in Section 2. Most studies focused on inverting the refractions. This is difficult, particularly when some of the objects are moving. Those moving objects are the actual interest, as they reveal animal, human, or machine activity. However, detecting and tracking these objects does not necessitate complete inversion of the medium and recovery of the scene. Deciding whether the motion of an image feature is solely due to refractive distortions or due to real object motion (compounded with distortion) is a *classification* problem. Given a good feature vector, classification can be obtained simply and computationally cheaply. This is the task of this paper. Discriminating consistently moving objects from a randomly moving background was studied in open air background subtraction [5], [26], [32], tracking [8], [51], and linear motion saliency [50]. In refractive distortions, however, moving objects are not *beside* a random background. Instead, the image of the moving object is prone to the same strong random distortions as static objects. In fact, motion due to distortions often dominates over object motion. In other words, motion *noise* can be larger than the motion *signal*. As our experiments show, interframe motion can span a significant portion of the field of view (FOV). This is much larger than typically considered in atmospheric turbulence [19], where local optical flow is used.

While the object and distortion motions are random and unknown, they are *mutually independent.* We exploit this fact. Being independent components, the location covariance due to object motion is *additive* to covariance stemming from medium dynamics. Hence, simple location covariance is a powerful classification feature. Characterizing the features corresponding to the class of static objects benefits from physics-based analysis of distortion statistics.

2 Related Work: Vision via a WAI

Approaches to vision through a WAI are as follows:

1. *Analysis of flat water distortion.* Distortions in oblique view through a flat WAI were thoroughly analyzed [10], [20], [36], [47]. Suiter et al. [43] introduce optical elements to the camera, to correct for flat WAI distortion. Our conclusion is that if a flat WAI is the only distortion source, then this distortion can be compensated for using stereo [10]. The remaining visual challenge is to overcome perturbations to the flat-WAI case. These are caused by random WAI waves.
2. *Physics-based recovery of a wavy WAI.* To compensate for WAI waves, some methods suggest detection of local flat WAI instances or dynamic reconstruction of the WAI. The former requires accumulation of sparse data from thousands of frames and a pulsed laser [27] synchronized to the camera. This complex and slow approach cannot handle moving objects. To recover the water geometry in real time, a theoretical work by Dolin [16] hypothesizes that the WAI slope can be inferred by its specular reflection[1] of a cloudless sky. Multicamera methods have also been applied to WAI recovery [15], [33], [37], [38], [49]. A critical review of passive approaches to recover the WAI is in [22]. Those studies focused on viewing downward into water. Looking up through the WAI to

1. Adato et al. [1] exploit the relative motion between a curved specular object and its environment to reconstruct its shape. This method is related to the recovery of specular surfaces from reflected distortions.

- *M. Alterman, Y.Y. Schechner, and J. Shamir are with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel. E-mail: amarina@tx.technion.ac.il.*
- *P. Perona is with the Department of Electrical Engineering, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125.*

Manuscript received 7 Aug. 2011; revised 25 July 2012; accepted 22 Aug. 2012; published online 4 Sept. 2012.
Recommended for acceptance by T. Pajdla.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2011-08-0532.
Digital Object Identifier no. 10.1109/TPAMI.2012.192.

Fig. 1. A frame of an airborne scene, taken by a submerged camcorder. The underwater scene is typically darker and smoother. It undergoes total internal reflection (TIR) by the water-air interface and appears in the lower part of the image. TIR is bounded by Snell's window where the airborne scene is seen. The strong random distortion changes in space and time. The task is to be able to sense objects that move and ignore static objects whose apparent motion is only due to the dynamics of the medium.

create a *virtual periscope* is theoretically proposed [30] based on a wave model constrained by self-occlusions. Polarization-based WAI recovery for visual correction was also theoretically proposed [39]. There are additional methods for measuring the water surface using reflection of structured light [3], [14]. They have not been tested as a tool for visual correction.

3. *Recovery of static objects via a wavy WAI.* Some methods bypass physical measurement of the WAI and directly aim to recover *static* objects, using video. For small distortions, Murase [34] uses optical flow to estimate the WAI normals, leading to image correction. There are patch-based versions of lucky imaging [17], [18], [48]. They accumulate a large number of frames, out which a best representative frame is selected per patch. Efros et al. [18] assume that the distribution of local image distortion is Gaussian. There, local patches are embedded in a feature space manifold. Patches close to the center of their manifold embedding are considered least distorted. Donate and Ribeiro [17] select a best set of low-distortion subregions from each video frame, using clustering. Then it combines the set to an undistorted image of the observed object. Wen et al. [48] select lucky patches aided by bispectral analysis to undistort the image. Tian and Narasimhan [44] fit a parametric distortion model to spatiotemporal data acquired with a fast camera looking straight downward into the water. The model is fitted to each frame to estimate the water surface and then to recover the underwater scene. In [35], multiple frames are registered to the temporal averaged image. Residual distortions are sparse and treated in a second stage. Tian and Narasimhan [45] seek image alignment given an undistorted template image. An undistorted template image is used to generate training samples for a known distortion model. Their novel pull-back operation utilizes these training images with known deformations to predict the parameters of the test image distortion. In contrast, our approach can work in absence of an undistorted template.

These approaches assume that the object is static while data is acquired. In our work, however, moving objects may exist in the FOV and are actually a focus of interest.

3 Motion Independence

Consider an airborne object at point $\mathbf{p}_a^{\text{lab}}$ in the lab coordinate system (see Fig. 3). The system's origin is at the center of projection of a submerged camera. If the WAI interface is completely flat, $\mathbf{p}_a^{\text{lab}}$ is projected to pixel $\mathbf{x}_{\text{flat}}$ in the detector plane, and the image is considered undistorted. We are only interested here in perturbations to that state: When the WAI is wavy, $\mathbf{p}_a^{\text{lab}}$ is projected[2] to pixel $\mathbf{x}$. Distortion thus takes the form of a shift,

$$\mathbf{s}^{\text{medium}} = \mathbf{x} - \mathbf{x}_{\text{flat}}, \tag{1}$$

in the image coordinates corresponding to the object point. Define the covariance matrix of $\mathbf{s}^{\text{medium}}$:

$$\mathbf{\Sigma}^{\text{medium}} = \begin{bmatrix} \left(\sigma_x^{\text{medium}}\right)^2 & \text{OD} \\ \text{OD} & \left(\sigma_y^{\text{medium}}\right)^2 \end{bmatrix}, \tag{2}$$

where the off diagonal (OD) elements are expectation over time t of $[s_x^{\text{medium}}(t) s_y^{\text{medium}}(t)]$. Here, σ_x^{medium} and σ_y^{medium} denote the respective standard deviations (STDs) in the respective image axes.

If the medium is calm, a moving object induces a shift of $\mathbf{s}^{\text{object}}$ pixels in its image coordinates. When both the object and the medium are dynamic, the projection of the object shifts by the vector

$$\mathbf{s}^{\text{observe}} = \mathbf{s}^{\text{medium}} + \mathbf{s}^{\text{object}}. \tag{3}$$

A static object is expressed by the special case $\mathbf{s}^{\text{object}} = 0$. Due to the random and fast kinetics of $\mathbf{s}^{\text{medium}}$, the shifts $\{\mathbf{s}^{\text{medium}}, \mathbf{s}^{\text{object}}\}$ are generally *independent* of each other. Due to the linear independence of the stochastic vectors $\mathbf{s}^{\text{medium}}$ and $\mathbf{s}^{\text{object}}$, the covariance matrix of $\mathbf{s}^{\text{observe}}$ is

$$\mathbf{\Sigma}^{\text{observe}} = \mathbf{\Sigma}^{\text{medium}} + \mathbf{\Sigma}^{\text{object}}, \tag{4}$$

where $\mathbf{\Sigma}^{\text{object}}$ is the covariance matrix of $\mathbf{s}^{\text{object}}$. Hence, object motion *increases* the diagonal elements of $\mathbf{\Sigma}^{\text{observe}}$. The diagonal elements of $\mathbf{\Sigma}^{\text{observe}}$ are simple *features* for classifying moving and stationary objects. Also, OD elements are affected by object motion, but as explained in Section 4.2, OD elements are often not as useful as the diagonal elements.

We do not consider defocus. Objects which are too close to the camera may be out of focus. Nevertheless, our analysis holds for the *chief ray*. For this ray, $\mathbf{x}$ and $\mathbf{x}_{\text{flat}}$, the motion statistics hold.

4 Stochastic Model

4.1 Distortion Statistics

The laboratory coordinate system is (ξ, η, z). Specifically, $\hat{\mathbf{z}} = (0, 0, 1)$ points to the zenith and (ξ, η) are lateral (horizontal) coordinates (see Fig. 3). Let the height of the WAI be z_{s} in this system. The WAI gradient at a given location is $\nabla z_{\text{s}} = [\partial_\xi z_{\text{s}}, \partial_\eta z_{\text{s}}]^{\text{T}}$, where T denotes transposition. Following Cox and Munk [13], the probability density distribution (PDF) of ∇z_{s} is

$$\text{PDF}(\nabla z_{\text{s}}) \sim \exp\left[-\frac{1}{2}(\nabla z_{\text{s}})^{\text{T}} \mathbf{\Sigma}_{\text{grad}}^{-1} \nabla z_{\text{s}}\right], \tag{5}$$

where $\mathbf{\Sigma}_{\text{grad}}$ is the covariance matrix of the WAI gradient. In nondirectional wind, $\mathbf{\Sigma}_{\text{grad}} = \sigma_{\text{grad}}^2 \mathbf{I}$, where $\mathbf{I}$ is the identity matrix and

$$\sigma_{\text{grad}}^2 = 0.003 + 0.0051 W. \tag{6}$$

Here, W is the wind speed in [meters/sec].

The WAI normal is related to ∇z_{s} by

$$\hat{\mathbf{N}} = \begin{bmatrix} -\nabla z_{\text{s}} \\ 1 \end{bmatrix} \left[(\nabla z_{\text{s}})^2 + 1\right]^{-1/2}. \tag{7}$$

2. An airborne point may project to multiple image points, through multiple paths. Our analysis considers only one image point per object point. We obtained successful classification results in real experiments despite this assumption.

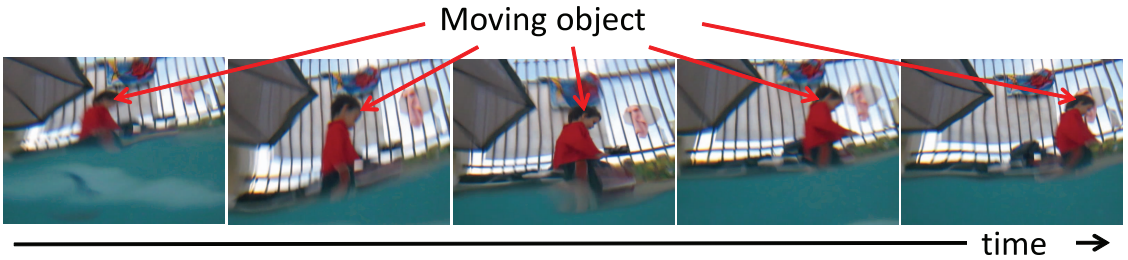

Fig. 2. Sample video frames. The entire scene is dynamically distorted. All objects undergo strong nonlinear and nonuniform stretching, shifting, and curving. There are objects having inherent motion, which is superimposed on the distortion.

Appendix A, which can be found in the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPAMI.2012.192, derives $\mathbf{s}^{\text{medium}}$ for a given WAI normal $\hat{\mathbf{N}}$ and height z_s. Following distribution (5), the surface normal $\hat{\mathbf{N}}$ and the medium distortion $\mathbf{s}^{\text{medium}} = (s_x^{\text{medium}}, s_y^{\text{medium}})$ are random in time. The WAI is smooth in space and time. Thus, there is spatiotemporal correlation of these random variables. It may be possible to exploit the correlations for more sophisticated models and classifier design. We process each point independently of others, using only temporal statistics. The distribution of the distortion can be drawn from the perturbation distribution of the medium (5), (6), and (7) and the distortion model. Details of the distortion model are in Appendix A, which is available in the online supplemental material.

Consider a typical setup. Let a camera have a focal length of 6.5 mm. The camera is submerged: $z_{\text{flat}} = 20$ cm. Underwater, it stares upward at 40 degrees from the zenith, corresponding to 31 degrees above the airborne horizon. Object point $\mathbf{p}_a^{\text{lab}}$ is at $z_a = 1.5$ meters above the WAI. WAI slopes are drawn from (5). Surface normals are computed using (7). In addition, the WAI height z_s is uniformly distributed around $z_{\text{flat}} \pm 2$ cm. We found, as expected, that the wave amplitude range has only a marginal effect on image distortion statistics, which are rather dominated by WAI gradients. Let $\mathbf{x}_{\text{flat}}$ be at the center of the FOV. The x-axis of the camera is perpendicular to the z-axis of the laboratory coordinate system. The simulated distribution of the distorted $\mathbf{x}$ is shown in Fig. 3.

Our experiments (see Section 7) were conducted in swimming pools. There, distortion statistics are determined by water depth, the size of the pool, and the fact that the waves were created as a result of swimmers rather than wind. Nevertheless, in this general study of distortion statistics, the sea surface statistical model of Cox and Munk (5) and (6) is taken as a representative. While this is an approximation of the experimental setup, it relies on a well-established theory and, as will be seen in Section 7, yields good experimental results in our task.

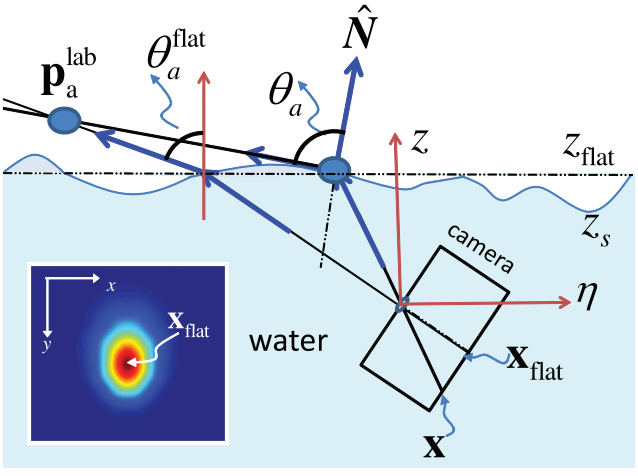

Fig. 3. Refraction at a wavy water surface. [INSERT] Numerically calculated distribution of $\mathbf{x}$ for a submerged camera close to the WAI, looking obliquely upward. The waves are distributed according to [13], corresponding to a nondirectional 4 knot wind. The WAI height z_s is uniformly distributed around $z_{\text{flat}} \pm 2$ cm. Here, the image size is 580×720, OD is null, $\sigma_x^{\text{medium}} \approx 130$ and $\sigma_y^{\text{medium}} \approx 170$ pixels

4.2 Observation Statistics

Recall that object motion increases the diagonal elements of $\boldsymbol{\Sigma}^{\text{observe}}$ (4). Thus, they are considered as simple *features* for classifying moving objects and stationary ones. In practice, $\boldsymbol{\Sigma}^{\text{observe}}$ is estimated based on the image data. Let $p = 1 \ldots N_{\text{points}}$ index object points. The observed (distorted) position of each object point is denoted by $\mathbf{x}_p = [x_p, y_p]^{\text{T}}$. It is tracked during N_{frames} frames. The mean position of each point is

$$\bar{\mathbf{x}}_p \equiv [\bar{x}_p, \bar{y}_p]^{\text{T}} = \frac{1}{N_{\text{frames}}} \sum_{t=1}^{N_{\text{frames}}} \mathbf{x}_p(t). \tag{8}$$

The estimated covariance matrix of each point is

$$\hat{\boldsymbol{\Sigma}}_p^{\text{observe}} = \frac{1}{N_{\text{frames}} - 1} \sum_{t=1}^{N_{\text{frames}}} [\mathbf{x}_p(t) - \bar{\mathbf{x}}_p][\mathbf{x}_p(t) - \bar{\mathbf{x}}_p]^{\text{T}}. \tag{9}$$

The feature vector of point p is comprised of the diagonal terms in $\hat{\boldsymbol{\Sigma}}_p^{\text{observe}}$, i.e., the estimated horizontal and vertical trajectory variances:

$$\mathbf{c}_p = \left[\hat{\sigma}_{\text{x},p}^2, \hat{\sigma}_{\text{y},p}^2\right]^{\text{T}}. \tag{10}$$

For simplicity, (10) does not use the OD terms of $\hat{\boldsymbol{\Sigma}}_p^{\text{observe}}$. As shown in Fig. 3, on the camera optical axis the OD terms of $\boldsymbol{\Sigma}^{\text{medium}}$ are null. Off axis, the OD terms are nonzero and small. Accounting for the known camera projection, the location distribution can be corrected for such off-axis effects. Hence, an OD term can theoretically be useful: Substantial OD values indicate a nonstatic object. However, this property can seldom be used: Almost always, objects of interest move horizontally (walking, driving, sailing, etc.) and sometimes they move vertically (animals drink from the water source, people load or unload gear, launch a spear in the water, etc.). Either of these motions has very small OD covariance. Appendix C, which is available in the online supplemental material, illustrates different motion types along with the corresponding covariance matrices. Nevertheless, our subsequent analysis can be applied to feature vectors comprising OD elements as well.

The elements of $\mathbf{c}_p$ are *not* the true variances of the trajectory, since they are not based on infinite N_{frames}. They are *estimates*, based on a small N_{frames}. As such, each $\mathbf{c}_p$ is a random sample from a distribution around the feature vector that would have been obtained had N_{frames} been infinite. The distribution width indicates the uncertainty of $\mathbf{c}_p$. Cochran's Theorem [9] states that if the observations are independent and normally distributed, their variance follows a scaled χ^2 distribution. Approximating the PDF of $\mathbf{s}^{\text{medium}}$ as Gaussian, the distribution of the elements of $\mathbf{c}_p$ is

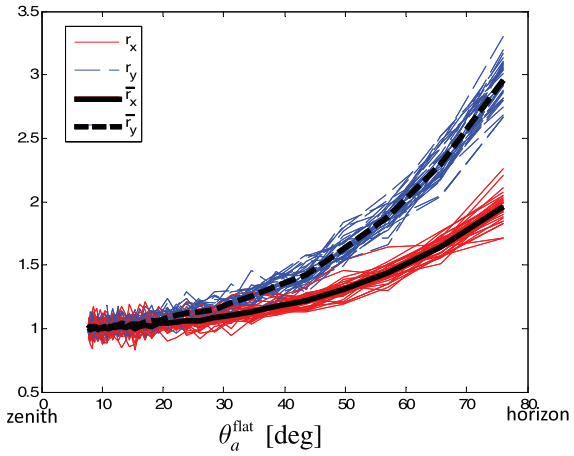

Fig. 4. Numerically calculated ratios r_x and r_y as a function of the airborne angle θ_a^{flat}. Each line corresponds to a different wind velocity W. Thick lines correspond to $\bar{r}_x$ and $\bar{r}_y$.

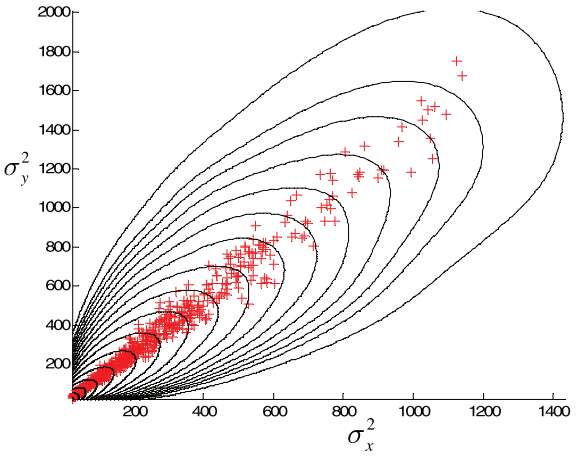

Fig. 5. Crosses mark a scatter plot of static objects in feature-space, theoretically derived for WAI statistics that correspond to wind conditions up to 2.5 knots. The plot is overlaid on a PDF. The contours mark level sets of the estimated PDF, and can be used as confidence intervals and thresholds.

$\sim \chi^2_{N_{\text{frames}}-1}$, following Cochran's Theorem. Thus, the expectation and STD of the first element in (10), $c_p(1) = \hat{\sigma}^2_{\text{x},p}$, are, respectively,

$$E[c_p(1)] = \sigma^2_{\text{x},p} \tag{11}$$

and

$$\text{STD}[c_p(1)] = \frac{\sqrt{2}\sigma^2_{\text{x},p}}{\sqrt{N_{\text{frames}}-1}} \approx \frac{\sqrt{2}c_p(1)}{\sqrt{N_{\text{frames}}-1}}, \tag{12}$$

with analogous expressions for $c_p(2) = \hat{\sigma}^2_{y,p}$.

The tracker may fluctuate. This induces noise in the observed position

$$\mathbf{x}_p = \mathbf{x}_p^{\text{true}} + \mathbf{n}^{\text{track}}, \tag{13}$$

where $\mathbf{x}_p^{\text{true}}$ is the true position and $\mathbf{n}^{\text{track}}$ is the tracker noise vector. Assuming independence of $\mathbf{x}_p^{\text{true}}$ and $\mathbf{n}^{\text{track}}$,

$$\hat{\mathbf{\Sigma}}^{\text{observe}} = \mathbf{\Sigma}^{\text{observe}}_{\text{true}} + \mathbf{\Sigma}^{\text{track}}. \tag{14}$$

Here, $\mathbf{\Sigma}^{\text{observe}}_{\text{true}}$ is the true covariance matrix of $\mathbf{s}^{\text{observe}}$ and $\mathbf{\Sigma}^{\text{track}}$ is the covariance matrix of $\mathbf{n}^{\text{track}}$. Hence, tracking noise biases the location variances (our classification features). This bias induces higher values in (12).

4.3 Dependence on the Viewing Angle

Let θ_a^{flat} be the LOS angle in air, relative to the zenith, when the WAI is flat (see Fig. 3). Section 4.1 described the derivation of $\mathbf{\Sigma}^{\text{medium}}$ for a given point in the FOV and wave conditions (wind speed W). Indeed, σ_x^{medium} and σ_y^{medium} are functions of θ_a^{flat} and W. This dependency can be assessed by simulations given θ_a^{flat} and wind conditions. As θ_a^{flat} increases, the pixel variance increases, particularly in the vertical pixel coordinate. Near grazing angles, even small perturbations in the WAI slope result in significant variance of $\mathbf{s}^{\text{medium}}$. In contrast, when looking straight upward (zenith), WAI waves yield only mild changes of the airborne LOS. This is expected, given Snell's law.

4.3.1 Wind Invariance

Define normalized STDs as the ratios

$$r_x = \frac{\sigma_x^{\text{medium}}}{\sigma_x^{\text{medium}}\left(\theta_a^{\text{flat}}=0^o\right)}, r_y = \frac{\sigma_y^{\text{medium}}}{\sigma_y^{\text{medium}}\left(\theta_a^{\text{flat}}=0^o\right)}. \tag{15}$$

Fig. 4 plots r_x and r_y for a range of W. It reveals that dependence on W is weak. This observation allows r_x and r_y to each be well approximated by a representative function, independent of W. We thus simply averaged either r_x and r_y, for a range of W. The respective results increase monotonically with θ_a^{flat}. They are used in Section 5 for feature calibration.

4.4 Physics-Based Feature Distribution

The distribution of static objects can be predicted, based on our theory. Suppose the inclination angle, focal length, FOV, and depth of the camera are all fixed and known. A major unknown is wind speed, which determines the WAI slope statistics. It also sets the WAI height distribution (which has just a marginal effect), according to the *Beaufort scale of wind force* [4].

Section 4.1 describes the derivation of $\mathbf{\Sigma}^{\text{medium}}$, for a given point in the FOV and wind speed. Hence, spreading points $\mathbf{x}_{\text{flat}}$ across the FOV and sampling wind according to some distribution yields a theoretical scatter plot in feature space, corresponding to static objects. As an example, we simulated vectors $\mathbf{c}_p$ in a range of the unknowns. The static object points $\mathbf{x}_{\text{flat}}$ correspond to views from 10 to 75 degrees above the airborne horizon. The wind speed is sampled uniformly from 0 to 2.5 knots. The scatter plot, shown as red crosses in Fig. 5, is mainly concentrated in the lower left part of the feature space. It extends to larger coordinates in feature space, sparsely. The extent depends on the maximum wind speed we allow.

Overall, the estimated joint PDF in feature space is illustrated in Fig. 5, using its level sets. Each level in Fig. 5 is also a confidence interval and a threshold: a certain percentage of static objects are expected to fall within the interval. Recall that due to tracking noise, observed features may be biased (13) and (14). We do not know the nature of $\mathbf{n}^{\text{track}}$ and the bias that it may cause. Thus, tracking noise is incorporated into the PDF by blurring it.

5 Camera and Feature Calibration

To analyze experimental movies, we first estimate two parameters that affect the pixel variance: the underwater effective focal length and the inclination angle of the camera. A submerged object of known length d^{obj} is captured by the submerged camera from a known distance D^{water}. The image corresponding to the projected object is d^{pix} pixels long. The effective focal length [in pixels] is $f_c = D^{\text{water}} \frac{d^{\text{pix}}}{d^{\text{obj}}}$.

The inclination (orientation) of the camera can be obtained by an accelerometer. Nowadays, smartphones integrate a camera, an orientation sensor, and a computer. Thus, it is technologically easy to know the inclination angle of a system without even looking through the camera. This in fact is already used in computer vision

work dealing with distortions [10]. Knowing the orientation of view also occurs in nature. Some fish sense and control their balance and orientation using otoliths [23] (similar to human sense of balance) or a swim-bladder [42].

In the absence of an orientation sensor, the inclination angle of the camera can be determined computationally by the location of Snell's window in the FOV (see Appendix B, which is available in the online supplemental material). As a result of knowing both the inclination and f_c, it is now possible to determine θ_a^{flat} for each camera pixel. This is detailed in Appendix B, which is available in the online supplemental material.

5.1 Feature Scaling

Section 4.3 shows that σ_x^{medium} and σ_y^{medium} increase with θ_a^{flat}. The known multiplicative factors $\bar{r}_x$ and $\bar{r}_y$ are plotted in Fig. 4. Furthermore, any image is an optical magnification: It scales linearly with f_c; thus, so do σ_x^{medium} and σ_y^{medium}. We seek to use a single classifier for all moving pixels, in all movies. Therefore, all features $\mathbf{c}_p$ (both during training and testing) are scaled to compensate for these systematic factors. Since generally $\bar{r}_x \neq \bar{r}_y$, scaling is different in each axis. The normalized features are

$$\tilde{\sigma}^2_{\text{x},p} = \frac{\hat{\sigma}^2_{\text{x},p}}{[\bar{r}_x(p) f_c]^2}, \quad \tilde{\sigma}^2_{\text{y},p} = \frac{\hat{\sigma}^2_{\text{y},p}}{[\bar{r}_y(p) f_c]^2}. \tag{16}$$

The factors $\bar{r}_x(p)$ and $\bar{r}_y(p)$ are individual to each feature p, as we explain. Equation (8) estimates the mean vertical pixel location $\bar{y}_p$. This approximates the vertical location of the tracked image feature, had the water been flat and the object static. Appendix B, which is available in the online supplemental material, derives the relation between the vertical pixel location $\bar{y}_p$ in flat water and $\theta_a^{\text{flat}}(p)$. Then, Fig. 4 yields $\bar{r}_x(p)$ and $\bar{r}_y(p)$.

6 Tracking

The empirical process relies on tracking of *interest points* (corners) corresponding to object points. If the object is severely defocused, some of its interest points disappear, hindering detection of movement. This problem is not specific to refractive distortions: Also in stable media, defocus blur hinders detection. A good practice is to focus the camera at a long distance, where most objects of interest are.

Tracking needs to work despite severe disturbances, one of which is strong dynamic distortions. Another disturbance is occasional loss of sight of the object due to several phenomena. First, dynamic refraction sometimes causes strong image blur, when the refraction changes very rapidly in space and time, leading to smear within the exposure time. The second phenomenon occurs near grazing airborn angles while the camera is submerged. In this case, WAI waves occasionally lead the LOS past the critical angle, yielding total internal reflection. Then, instead of acquiring the airborne object, the camera briefly captures a reflection of the underwater scene. Loss of sight of an object resembles occlusion, but here the disturbance disappears as fast as it appears, or moves nonrigidly.

Sudden disappearance and then reappearance of the object makes it difficult to employ parametric [2] image rectification. The natural scene complications call for representation of motion and trajectory, which is more mid-level than optical flow of image irradiance. We thus opt for feature tracking since tracking has been developed in the literature to sustain geometric and radiometric distortions as well as occlusions. We have no particular preference for any tracker. We performed a quantitative comparison of several trackers. Tracking 260 interest points in different movies and WAI wave conditions yielded a 77 percent success rate for a large displacement optical flow (LDOF) tracker [6], [7], 63 percent for a tracking-learning-detection (TLD) [24] approach, and 45 percent for the Kernel-based tracker [12]. Thus, we selected the LDOF tracker for our experiments. In this approach, correspondences from descriptor matching are integrated into a variational optical flow model.

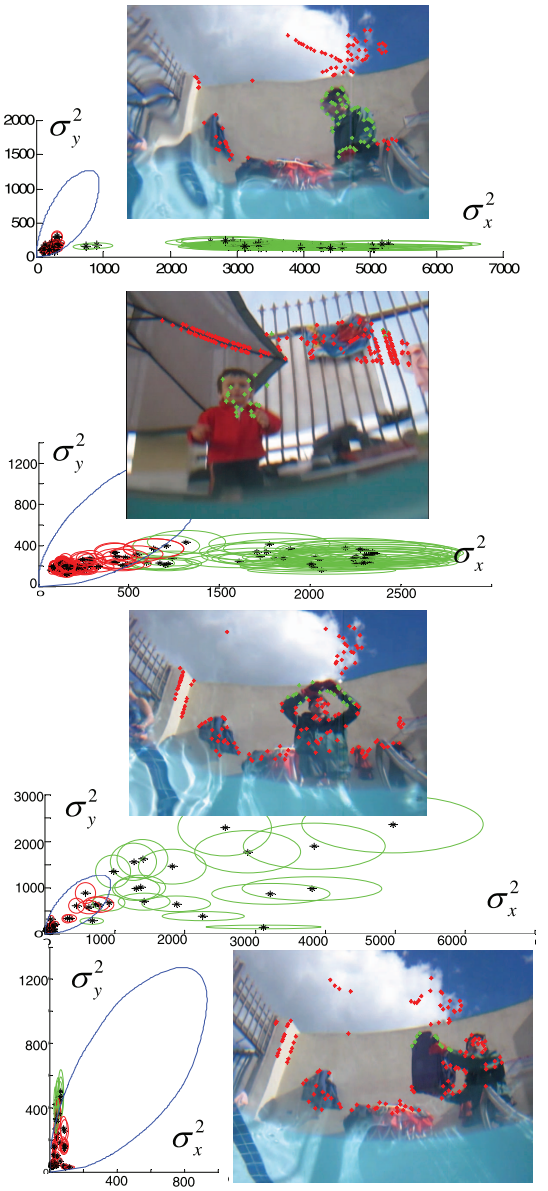

Fig. 6. The fixed theoretical classifier contour in feature space, and scatter plots of several test-sets. Upper: Horizontally walking subject. Lower: Vertically moving object.

7 Classification Experiments

We conducted experiments in a swimming pool using a video-recording camera (Canon PowerShot D10 mounted on a tripod at various focal length settings.

Each movie is several seconds long. We used $N_{\text{frames}} \approx 30$-$40$ in a temporally sliding window to estimate the trajectory statistics each time. In other words, about a second of acquisition was allocated to the task of directly detecting moving objects during dynamic refraction. Interest points were tracked. Fig. 6 shows examples extracted from real movies. Each feature in the scatter plot is plotted as an ellipse: Its center is at $\mathbf{c}_p$, while its widths are set according to (12). Red ellipses represent static object points, while green ellipses represent dynamic object points. The

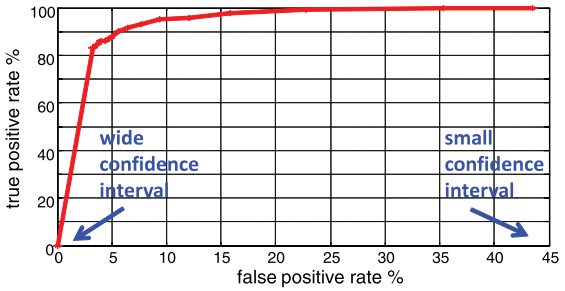

Fig. 7. ROC curve for different confidence interval threshold levels. A low false positive rate corresponds to wide static confidence interval, while a high false positive rate corresponds to a small interval.

corresponding red and green points are superimposed on the sample movie frames.

For classification, a decision contour in this feature space is set based on the feature distribution of static objects. This distribution is assessed from the theoretical simulation, as was explained in Section 4.4. Each of the contours in Fig. 5 can be selected as the classifier. A wider contour corresponds to fewer false-positives (static objects falsely classified as moving) in testing and more false-negatives (moving objects falsely classified as static). This is summarized by an ROC [52] curve (Fig. 7).

The classifier was based on thresholding on a confidence interval of the theoretical PDF. The threshold corresponds to ~ 85 percent true positive rate in Fig. 7. The classifying curve is shown as a blue line in Fig. 6. Due to the fixed classifier setting, some examples yield some false positives or negatives. Tuning the classifier threshold eliminates many of these errors.

Classification failure can occur for dynamic objects that have very small movements during the acquisition time: Such objects may be classified as static because their feature-vectors fall inside the static-object threshold curve. Feature-vectors of such moving objects resemble the feature-vectors of water fluctuations. So, the objects cannot be separated solely based on variance features.

7.1 Distortion Correction for Motion Detection

Multiple-image techniques for image rectification such as [17], [18], [35], [44] rely on a video sequence of a static scene. Thus, they are not suitable for the situation having moving objects in the scene. It may be possible to apply single-image distortion correction per video frame. The approach in [45] is such a single-image approach; however, it requires an undistorted template image. There are many practical cases, in nature, in photography from submarines, and in our experiments, where an undistorted template image of the scene does not exist. The method of [45] was not designed to work in such situations, in contrast to our approach.

Let us assume that, in training, an undistorted reference image of the scene exists while it is static. Then in testing, static pixels can potentially be registered to the undistorted template using the method of [45]. Pixels that cannot be registered to the template should be classified as moving. In our experiments, the required undistorted template image is unavailable. Hence, we set the first frame (with no moving objects) to be a template frame. After obtaining a rectified video using an existing implementation [46], we subtracted each original video frame from the template image, to find inconsistences (moving objects). As shown in Fig. 8, there are many false positives (static objects classified as moving). This result may not do justice to the approach of [45], as it had not been designated to work in our situation. Moreover, note that the approach of [45] is dense (all pixels are classified), whereas our approach is sparse (only tracked points are classified). Thus, both techniques are not comparable in terms of absolute number of false positives. Basically, this approach is promising. Thus, further research into the approach of [45] may improve it to yield better image rectification and thus better classification results.

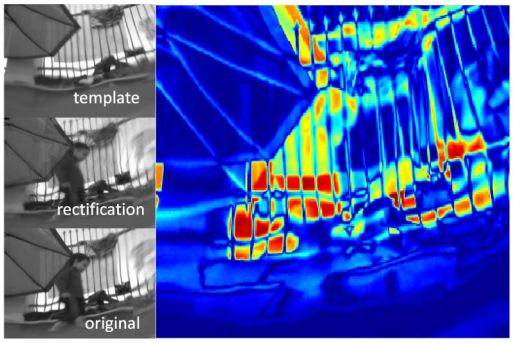

Fig. 8. Classification experiment utilizing the method in [45]. The differences between the template and the rectification images. Red presents high difference values, interpreted as moving objects. Blue presents low difference values, interpreted as static objects. Many false positives are observed.

8 Discussion

To the best of our knowledge, this work is the first to raise and characterize the problem of detecting moving objects in highly random refractive distortions. The work models the statistics of motion under dynamic distortions. We evaluate both the statistics and its error theoretically and numerically. Features corresponding to the class of static objects were characterized, aided by the physics-based analysis of distortion statistics. A conclusion is that simple location covariance is a powerful classification feature. Tracked points whose feature vector falls outside the thresholded distribution of static objects can be classified as being on a moving object. The rate of false negatives and positives depends on the threshold, as in any classification problem.

We found the tracker we used to be robust enough to cope with a high degree of image distortion in our practical experiments where the resulting trajectories led to highly successful classification results. However, there are extreme situations in which a tracker fails and, as a result, a classification error can occur. In very harsh fluctuations of the medium, an instantaneous video frame may become very blurred due to very fast medium motion. This motion blur may sometimes cause the tracker to pick up on a different target as a result of smearing between nearby objects. Perhaps this problem can be overcome by excluding extremely motion-blurred frames from tracking, for example by blur detection [28]. In addition, the rapidly varying shape of water waves may cause an airborne point to be projected to multiple points on the detector plane. This also causes the image to be blurred and thus degrades the tracker performance.

A classification failure is also likely to occur for very slow moving objects, aliasing as static in the temporal window. In nature, this situation resembles predators sneaking up on prey. Feature-vectors of such slowly moving objects are close to the class of actual static objects, leading to detection error.

We believe that our approach can also be useful in milder distortions, as occurs when looking downward into water and in atmospheric turbulence. In short, the approach is as follows: First use the trajectory covariance as a feature vector. The property of independent motion components and increase of variance holds in these cases as well. Then, characterize the distribution of static objects, either by physics-based calculations or by training. Then a simple classifier is at hand.

Acknowledgments

The authors would like to thank the anonymous reviewers for their useful comments. They thank Tomer Schechner, Yotam Michael, Amit Aides, and Amit Oved for their participation and help in the experiments. Experiments were conducted in the swimming pool facilities of Caltech, and the authors are grateful for this. They appreciate the occasional visitors in the Caltech pool and spa for

useful discussions. They also thank Ohad Ben-Shahar for useful discussions about fish orientation. They thank Yuandong Tian and Srinivasa G. Narasimhan for sharing their code for distortion estimation. Yoav Schechner is a Landau Fellow—supported by the Taub Foundation. Marina Alterman's research is supported by the R.L. Kohns Eye Research Fund. This work relates to US Department of the Navy Grant N62909-10-1-4056 issued by the Office of Naval Research Global. The United States Government has a royalty-free license throughout the world in all copyrightable material contained herein.

REFERENCES

[1] Y. Adato, Y. Vasilyev, T. Zickler, and O. Ben-Shahar, "Shape from Specular Flow," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 32, no. 11, pp. 2054-2070, Nov. 2010.

[2] S. Agarwal, S. Mallick, D. Kriegman, and S. Belongie, "On Refractive Optical Flow," *Proc. European Conf. Computer Vision,* 2004.

[3] S. Baglio, C. Faraci, and E. Foti, "Structured Light Approach for Measuring Sea Ripple Characteristics," *Proc. Conf. OCEANS,* vol. 1, pp. 449-453, 1998.

[4] http://whale.wheelock.edu/whalenet-stuff/beaufort.html, 2012.

[5] O. Boiman and M. Irani, "Detecting Irregularities in Images and in Video," *Int'l J. Computer Vision,* vol. 74, pp. 17-31, 2007.

[6] T. Brox and J. Malik, "Large Displacement Optical Flow: Descriptor Matching in Variational Motion Estimation," *IEEE Trans Pattern Analysis and Machine Intelligence,* vol. 33, no. 3, pp. 500-513, Mar. 2011.

[7] T. Brox, "Large Displacement Optical Flow Code,$rdquo; http://132.230.167.110/people/brox/resources/pami2010Matlab.zip, 2012.

[8] A. Bugeau and P. Pérez, "Detection and Segmentation of Moving Objects in Highly Dynamic Scenes," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2007.

[9] W. Cochran, "The Distribution of Quadratic Forms in a Normal System, with Applications to the Analysis of Covariance," *Proc. Cambridge Philosophical Soc.,* vol. 30, pp. 178-191, 1934.

[10] Y.J. Chang and T. Chen, "Multi-View 3D Reconstruction for Scenes under the Refractive Plane with Known Vertical Direction," *Proc. IEEE Int'l Conf. Computer Vision,* 2011.

[11] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach toward Feature Space Analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 5, pp. 603-619, May 2002.

[12] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-Based Object Tracking," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 5, pp. 564-575, May 2003.

[13] C. Cox and W. Munk, "Statistics of the Sea Surface Derived from Sun Glitter," *J. Marine Research,* vol. 13, pp. 198-227, 1954.

[14] D. Dabiri and M. Gharib, "Interaction of a Shear Layer with a Free Surface," *Proc. Int'l Symp. Flow Visualization,* 2000.

[15] Y. Ding, F. Li, Y. Ji, and J. Yu, "Dynamic 3D Fluid Surface Acquisition Using a Camera Array," *Proc. IEEE Int'l Conf. Computer Vision,* 2011.

[16] L.S. Dolin, A.G. Luchinin, V.I. Titov, and D.G. Turlaev, "Correcting Images of Underwater Objects Distorted by Sea Surface Roughness," *Proc. SPIE,* vol. 6615, p. 66150K, 2007.

[17] A. Donate and E. Ribeiro, "Improved Reconstruction of Images Distorted by Water Waves," *Proc. Int'l Conf. Computer Vision Theory and Applications,* pp. 228-235, 2006.

[18] A. Efros, V. Isler, J. Shi, and M. Visontai, "Seeing through Water," *Neural Information Processing Systems,* vol. 17, pp. 393-400, 2004.

[19] B. Fishbain, L. Yaroslavsky, and I. Ideses, "Real-Time Stabilization of Long Range Observation System Turbulent Video," *J. Real-Time Image Processing,* vol. 2, pp. 11-22, 2007.

[20] G. Horváth and D. Varjú, "On the Structure of the Aerial Visual Field of Aquatic Animals Distorted by Refraction," *Bull. Math. Biology,* vol. 53, pp. 425-441, 1991.

[21] I. Ihrke, K.N. Kutulakos, H.P. Lensch, M.A. Magnor, and W. Heidrich, "State of the Art in Transparent and Specular Object Reconstruction," *Proc. Eurographics,* 2008.

[22] B. Jähne, J. Klinke, and S. Waas, "Imaging of short Ocean Wind Waves: A Critical Theoretical Review," *J. Optical Soc. Am.,* vol. 11, pp. 2197-2209, 1994.

[23] H. Kleerekoper and T. Malar, *Orientation through Sound in Fishes in Hearing Mechanisms in Vertebrates,* 1968.

[24] Z. Kalal, J. Matas, and K. Mikolajczyk, "Online Learning of Robust Object Detectors during Unstable Tracking," *Proc. On-Line Learning for Computer Vision Workshop,* 2009.

[25] G. Katzir and N. Intrator, "Striking of Underwater Prey by a Reef Heron, Egretta Gularis Schistacea," *J. Computational Physics A,* vol. 160, pp. 517-523, 1987.

[26] T. Ko, S. Soatto, and D. Estrin, "Warping Background Subtraction," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2010.

[27] I.M. Levin, V.V. Savchenko, and V.J. Osadchy, "Correction of an Image Distorted by a Wavy Water Surface: Laboratory Experiment," *Applied Optics,* vol. 47, pp. 6650-6655, 2008.

[28] R. Liu, Z. Li, and J. Jia, "Image Partial Blur Detection and Classification," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 1-8, 2008.

[29] B. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *Proc. Seventh Int'l Joint Conf. Artificial Intelligence,* vol. 81, pp. 674-679, 1981.

[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 Complex Media,* vol. 16, no. 4, pp. 521-530, 2006.

[31] D. Miyazaki and K. Ikeuchi, "Shape Estimation of Transparent Objects by Using Inverse Polarization Ray Tracing," *IEEE Trans Pattern Analysis and Machine Intelligence,* vol. 29, no. 11, pp. 2018-2030, Nov. 2007.

[32] A. Monnet, A. Mittal, N. Paragios, and V. Ramesh, "Background Modeling and Subtraction of Dynamic Scenes," *Proc. IEEE Int'l Conf. Computer Vision,* 2003.

[33] N.J.W. Morris and K.N. Kutulakos, "Dynamic refraction stereo," *Proc. IEEE Int'l Conf. Computer Vision,* 2005.

[34] H. Murase, "Surface Shape Reconstruction of a Nonrigid Transport Object Using Refraction and Motion," *IEEE Trans Pattern Analysis and Machine Intelligence,* vol. 14, no. 10, pp. 1045-1052, Oct. 1992.

[35] O. Oreifej, G. Shu, T. Pace, and M. Shah, "A Two-Stage Reconstruction Approach for Seeing through Water," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2011.

[36] H. Saito, H. Kawamura, and M. Nakajima, "3D Shape Measurement of Underwater Objects Using Motion Stereo," *Proc. IEEE 21st Int'l Conf. Industrial Electronics, Control, and Instrumentation,* pp. 1231-1235, 2002.

[37] M.S. Schmalz, "Integration of Stereophotogrammetry with Image Restoration Models for Distortion-Tolerant Viewing through the Sea Surface," *Proc. SPIE,* vol. 1943, pp. 115-128, 1993.

[38] H. Schultz, "Shape Reconstruction from Multiple Images of the Ocean Surface," *Surface, Photogrammetric Eng. and Remote Sensing,* vol. 62, pp. 93-99, 1994.

[39] H. Schultz and A. Corrada-Emmanuel, "System and Method for Imaging through an Irregular Water Surface," US Patent 7,630,077, 2007.

[40] S. Schuster, S. Rossel, A. Schmidtmann, I. Jager, and J. Poralla, "Archer Fish Learn to Compensate for Complex Optical Distortions to Determine the Absolute Size of Their Aerial Prey," *Current Biology,* vol. 14, pp. 1565-1568, 2004.

[41] Q. Shan, B. Curless, and T. Kohno, "Seeing through Obscure Glass," *Proc. European Conf. Computer Vision,* 2010.

[42] J.B. Steen, "The Swim Bladder as a Hydrostatic Organ," *Fish Physiology: The Nervous System, Circulation, and Respiration,* vol. 4, 1970.

[43] 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," *Proc. Conf. OCEANS,* 2008.

[44] Y. Tian and S.G. Narasimhan, "Seeing through Water: Image Restoration Using Model-Based Tracking," *Proc. IEEE Int'l Conf. Computer Vision,* 2009.

[45] Y. Tian and S.G. Narasimhan, "A Globally Optimal Data-Driven Approach for Image Distortion Estimation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2010.

[46] Y. Tian and S.G. Narasimhan, "A Globally Optimal Data-Driven Approach for Image Distortion Estimation Code," http://www.cs.cmu.edu/ ILIM/projects/IM/globalopt/research_globalopt.html, 2012.

[47] T. Treibitz, Y.Y. Schechner, and H. Singh, "Flat Refractive Geometry," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2008.

[48] Z. Wen, A. Lambert, D. Fraser, and H. Li, "Bispectral Analysis and Recovery of Images Distorted by a Moving Water Surface," *Applied Optics,* vol. 49, pp. 6376-6384, 2010.

[49] R. Westaway, S. Lane, and D. Hicks, "Remote Sensing of Clear-Water, Shallow, Gravel-Bed Rivers Using Digital Photogrammetry," *Photogrammetric Eng. and Remote Sensing,* vol. 67, pp. 1271-1282, 2001.

[50] L. Wixson, "Detecting Salient Motion by Accumulating Directionally-Consistent Flow," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 8, pp. 774-780, Aug. 2002.

[51] Q. Zhu, S. Avidan, and K. Cheng, "Learning a Sparse, Corner-Based Representation for Time-Varying Background Modelling," *Proc. IEEE Int'l Conf. Computer Vision,* 2005.

[52] M.H. Zweig and G. Campbell, "Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine," *Clinical Chemistry,* vol. 39, no. 561, 1993.