

# CauStereo: Range from light in nature

Yohay Swirski,<sup>1,\*</sup> Yoav Y. Schechner,<sup>1</sup> Ben Herzberg,<sup>1</sup> and Shahriar Negahdaripour<sup>2</sup>

<sup>1</sup>Department of Electrical Engineering, Technion—Israel Inst. Technology, Haifa 32000, Israel

<sup>2</sup>Electrical and Computer Engineering Department, University of Miami, Coral Gables, Florida 33124-0640, USA

\*Corresponding author: yohays@tx.technion.ac.il

Received 23 May 2011; accepted 19 July 2011;  
posted 16 August 2011 (Doc. ID 147987); published 26 September 2011

Underwater, natural illumination typically varies strongly temporally and spatially. The reason is that waves on the water surface refract light into the water in a spatiotemporally varying manner. The resulting underwater illumination field forms a caustic network and is known as flicker. This work shows that caustics can be useful for stereoscopic vision, naturally leading to range mapping of the scene. Range triangulation by stereoscopic vision requires the determination of correspondence between image points in different viewpoints, which is often a difficult problem. We show that the spatiotemporal caustic pattern very effectively establishes stereo correspondences. Thus, we term the use of this effect as CauStereo. The temporal radiance variations due to flicker are unique to each object point, thus disambiguating the correspondence, with very simple calculations. Theoretical limitations of the method are analyzed using ray-tracing simulations. The method is demonstrated by underwater *in situ* experiments. © 2011 Optical Society of America

OCIS codes: 010.7295, 010.7340, 330.1400.

## 1. Introduction

Natural underwater environments exhibit several optical effects in typical lengths that are somewhat comparable to human and animal lengths. These include scattering, absorption, and spatiotemporally varying random scene irradiance [1–4]. The latter phenomenon is sometimes termed sunlight flicker [5] or caustic networks [6]. An example is shown in Fig. 1(a). This phenomenon is caused by refraction of sunlight through the wavy water surface [6] (Upwelling radiance through the surface was studied in [7]).

The mentioned optical phenomena strongly affect underwater vision. Analysis of visual signals in this environment is important for understanding biological vision of marine animals [8–12]. It is also required in man-made systems handling oceanic engineering tasks [13–16] and is useful in archaeological mapping [17]. Hence, computer vision methods have been developed for this environment, as

adaptations of open-air computer vision methods. Some computer vision methods try to tackle poor visibility conditions [18,19], which often exist [1] underwater.

In the context of vision, the temporal inconsistency of caustic networks has typically been considered mainly as a significant disturbance. Thus, attempts were made to reduce this effect by image postprocessing [5,20]. However, in this work we show that natural flicker can actually be informative and useful for vision.

An important visual task is estimation of the scene's range map. This is commonly achieved using a stereoscopic setup. Two cameras (or eyes), image the scene from two different viewpoints. Then, image features (e.g., pixels or blocks) in the two viewpoints are mutually matched. Once a correspondence between image features is established, the range to the corresponding scene region is triangulated. However, establishing correspondence is often difficult, prone to ambiguities and errors.

This work shows that spatiotemporal variations of caustic networks solve this problem rather simply. Thus, we term the use of this effect as CauStereo.

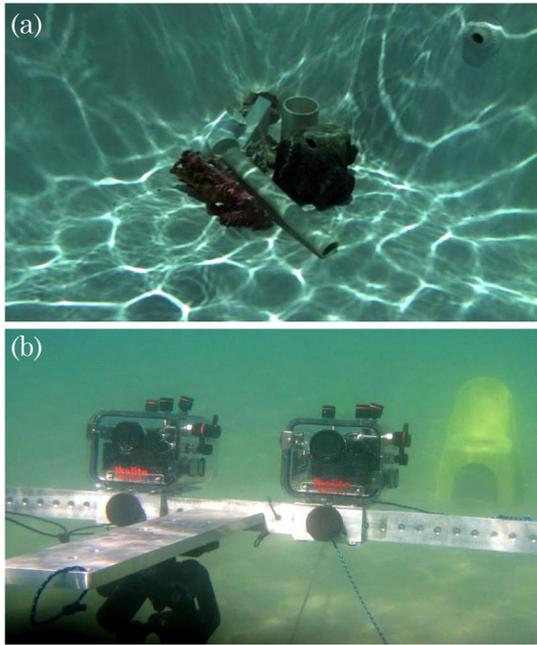


Fig. 1. (Color online) (a) Sunlight flicker irradiating a scene in a pool. (b) Underwater stereoscopic video setup in the Mediterranean. Taken from [21] with permission ©2009 IEEE.

The temporal radiance variations due to flicker are unique to each object point, thus disambiguating the correspondence. This is demonstrated by underwater *in situ* experiments. The model underlying the approach includes simplifying assumptions. The limitation of the main assumption is therefore studied as well. This is done using ray-tracing simulations. Partial results appeared in [21,22].

## 2. Theoretical Background

### A. Stereo and Correspondence

As shown in Fig. 1(b), a submerged scene is imaged by two cameras from different viewpoints. The centers of projection of the two cameras are mutually separated by a baseline  $b$ , as illustrated in Fig. 2.

Denote the left camera by  $L$ . We align the global coordinate system to this camera, i.e., the position

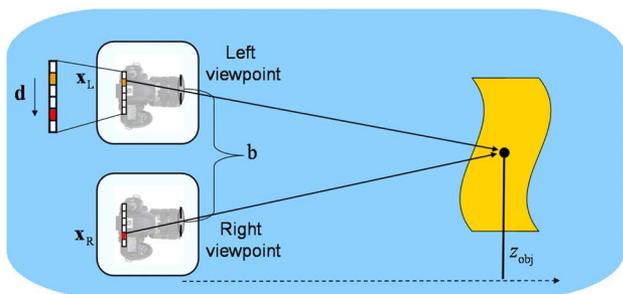


Fig. 2. (Color online) Underwater stereoscopic pair consisting of two cameras inside housings. The cameras are separated by baseline  $b$ . An object point at distance  $Z_{obj}$  is projected into the cameras at  $x_L$  and  $x_R$ . If correspondence between the cameras is known for the object point, it is possible to triangulate and calculate  $Z_{obj}$  out of the disparity  $d$  between the viewpoints.

of a scene point is uniquely defined by the left spatial coordinate vector  $x_L$  and the distance of an object point from the stereo rig  $Z_{obj}$ . The right camera is denoted by  $R$ . The object corresponding to  $(x_L, Z_{obj})$  in the left camera is projected to pixel  $x_R$  in the right camera.

Assume for a moment that there is only a single point source of light in the field of view, and no light is scattered by the volume or surface. Then, each of the images is dark, except for a single point in each ( $x_L$  and  $x_R$ , respectively). In this case, correspondence between the bright image points is trivially established. Then, the point source location can be triangulated by back-projecting  $x_L$  and  $x_R$  into the volume, along the rays illustrated in Fig. 2. Specifically, the disparity vector of the two points is

$$d = x_R - x_L. \quad (1)$$

Let  $d = \|d\|$ . For on-axis object points

$$Z_{obj} = bF/d, \quad (2)$$

where  $F$  is the distance between the camera exit pupil and its detector array. Following Eq. (2), the range map of the scene,  $Z_{obj}(x_L)$  is indicated by the inverse of the disparity map  $d(x_L)$ .

Of course, natural underwater scenes do not contain a single radiation point source: the entire field of view acquires light scattered and reflected in the scene. This makes matching of  $x_R$  to any  $x_L$  somewhat more difficult, in what is known to be the correspondence problem. Elaborate algorithms for solving this problem have been developed during the past decades [23] by the computer vision community. Essentially, they rely on spatial correlation between viewpoints: a spatial block of pixels in the left image is used as a template, for which a match is searched in the right image. Because of the extended spatial support, such methods face resolution difficulties and frequently fail near distance edges.

This paper shows that the natural phenomenon of spatiotemporal caustics enables an easy and accurate solution to the stereo correspondence problem. Thanks to this effect, there is no need to use a wide spatial support for matching wherever temporal light variations exist: all calculations can be point-wise. This bypasses the problems associated with methods that rely on spatial content.

### B. Underwater Flickering Caustics

Consider a short period of time. Generally, during this period, the water surface is wavy [24–26]. Therefore, the refraction of sunlight through the water surface creates inhomogeneous irradiance underwater. The resulting irradiance pattern is random. When light reaches the sea floor or other underwater objects, it forms bright regions in a caustic network [6]. Because of the natural motion and evolution of water surface waves, this light pattern changes in time. Thus, it is sometimes referred to as sunlight

flicker [5]. This phenomenon occurs only in regions that are illuminated by direct sunlight. In this paper, the underwater irradiance is denoted by  $I^{\text{lighting}}(\mathbf{x}_L, Z_{\text{obj}}, m)$ , where  $m$  is the temporal frame index.

How does sunlight flicker change with depth? As reported in [20], the problem was studied in [1]. The relative power of irradiance fluctuations decays exponentially with underwater depth, at a rate that depends on the water turbidity: in clearer water, sunlight flicker remains significant deeper than in turbid water. Thus, as rule of thumb, flicker is not negligible at depths of meters to  $\sim 20$  meters, depending on the water clarity. Furthermore, the temporal changes of flicker slow down with depth [1,20]. Hence, longer periods are needed in order to sense and make use of flicker, as objects are submerged deeper.

### C. Signal and Backscatter

Let the object radiance at a point be  $I^{\text{obj}}(\mathbf{x})$ . Because of attenuation in the water, the signal originating from this object [18] is

$$S(\mathbf{x}) = I^{\text{obj}}(\mathbf{x}) \exp\{-\eta[Z_{\text{obj}}(\mathbf{x}) - Z_h]\}, \quad (3)$$

where  $\eta$  is the attenuation coefficient of the water. Here  $Z_{\text{obj}}(\mathbf{x})$  is the distance to the object at  $\mathbf{x}$  and  $Z_h$  is the distance between the entrance pupil of the camera and the watertight housing in which the camera resides.

In addition to the object signal, the camera also senses path radiance, which is caused by ambient illumination scattered into the line of sight (LOS) by the water. This component is also termed backscatter [27], and is denoted by  $B(\mathbf{x})$ . It is given [18] by an integral over the LOS:

$$B(\mathbf{x}) = \int_{\tilde{z}=Z_h}^{Z_{\text{obj}}(\mathbf{x})} I^{\text{lighting}}(\tilde{z}) p(\theta) \exp(-\eta\tilde{z}) d\text{LOS}_{\mathbf{x}}, \quad (4)$$

where  $d\text{LOS}_{\mathbf{x}}$  is a differential along the LOS. Here  $p$  is the phase function and  $\theta$  is the lighting angle relative to the LOS, at a scatterer on the LOS. Equation (4) can be approximated [18,28] as

$$B(\mathbf{x}) = B_{\infty}(1 - \exp\{-\eta[Z_{\text{obj}}(\mathbf{x}) - Z_h]\}), \quad (5)$$

where  $B_{\infty}$  is the path radiance in a LOS reaching infinity. Overall, the radiance measured by the camera is

$$I(\mathbf{x}) = S(\mathbf{x}) + B(\mathbf{x}). \quad (6)$$

## 3. Underwater Stereoscopic Model

The image formation model is simple. It combines the spatiotemporal irradiance field described in Section 2.B with the LOS effects described in Section 2.C. We adapt the formulation to stereo.

In addition, we employ approximations that simplify the model and the resulting shape recovery.

### A. Flicker Signal

In the left camera, the signal corresponding to Eq. (3) is

$$S_L(\mathbf{x}_L, m) = I^{\text{lighting}}(\mathbf{x}_L, Z_{\text{obj}}, m) r_L(\mathbf{x}_L) \times \exp\{-\eta[Z_{\text{obj}}(\mathbf{x}_L) - Z_h]\}, \quad (7)$$

where  $r_L$  denotes the reflectance coefficient of the object toward the left camera. The viewpoints of the two cameras are different, separated by a baseline. This has two consequences. First, the distance between the object and the left camera is different than the distance of the object from the right camera. This affects the signal attenuation. Second, the direction from an object patch to the left camera is different than the direction to the right camera. This may affect the reflectance coefficient  $r$ .

For reasons that we detail below, we assume that these differences do not have significant consequences on the eventual range recovery. Overall we use

**Approximation 1:** At both cameras, the signal is attenuated in the same manner,  $\approx \exp\{-\eta[Z_{\text{obj}}(\mathbf{x}_L) - Z_h]\}$ . Differences of attenuation have no significance.

**Approximation 2:** The reflectance coefficients are the same for both cameras:  $r_L = r_R$ . Differences between these coefficients have no significance.

Suppose for the moment that the attenuation to the right camera is significantly different than the left, by a factor  $f_1(\mathbf{x}_R)$ . Furthermore, suppose the corresponding reflectance coefficients are different by a factor  $f_2(\mathbf{x}_R)$ . Then, the signal in the right camera is

$$S_R(\mathbf{x}_R, m) = I^{\text{lighting}}(\mathbf{x}_L, Z_{\text{obj}}, m) r_L(\mathbf{x}_L) \exp\{-\eta[Z_{\text{obj}}(\mathbf{x}_L) - Z_h]\} f_1(\mathbf{x}_R) f_2(\mathbf{x}_R) \\ = S_L(\mathbf{x}_R, m) f_1(\mathbf{x}_R) f_2(\mathbf{x}_R). \quad (8)$$

The factors  $f_1$  and  $f_2$  are a function of space, but are temporally invariant. Hence, they are canceled out if matching between  $\mathbf{x}_L$  and  $\mathbf{x}_R$  is established by normalized temporal correlation of the signals.

These approximations are supported by other reasons. The distance difference to the two cameras is typically much lower than  $b$ , since the baseline is approximately perpendicular to the optical axis, and  $b \ll Z_{\text{obj}}(\mathbf{x}_L)$ . Anyway, the distance difference is bounded by  $b$ . Thus,  $f_1(\mathbf{x}_R)$  is bounded by  $\exp(-\eta b)$ . Typically,  $\eta b \ll 1$ . For example, in water having visibility distance ( $1/\eta$ ) of 10 m and baseline of 30 cm,  $\exp(-\eta b) \approx 0.97$ . Hence,  $f_1(\mathbf{x}_R) \approx 1$ . The assumption  $f_2(\mathbf{x}_R) \approx 1$  is common in the stereo matching literature, and is known as the brightness constraint. It holds if in the two viewpoints the object appears Lambertian. In water, objects often satisfy this condition better than in air [29]. Recapping, following the above approximations in the context of Eq. (8), signals in the two cameras are related by

$$S_L(\mathbf{x}_L, m) \approx S_R(\mathbf{x}_R, m) \quad \forall m. \quad (9)$$

## B. Backscatter

Backscatter under flickering illumination depends both on time and viewpoint. Rewriting Eq. (4) for the left camera,

$$B_L(\mathbf{x}_L, m) = \int_{\tilde{z}=Z_h}^{Z_{\text{obj}}(\mathbf{x}_L)} I^{\text{lighting}}(\tilde{z}, m) p(\theta_L) \exp(-\eta\tilde{z}) d\text{LOS}_{\mathbf{x}_L}. \quad (10)$$

Is the spatiotemporal variation of the integral significant? Note that flicker hardly changes the overall energy irradiated into the water. At each frame, the caustic pattern essentially changes the spatial distribution of this energy. In this random spatial pattern, some regions on the LOS have higher irradiance than others. However, different regions have different weight in Eq. (10): a closer water voxel contributes more to the overall backscatter than a distant voxel. If the caustic distributes much of its energy at nearby voxels, the backscatter can be significantly higher than if this energy is focused only at very distant voxels. So, an important question is, how wide are the features of the caustic pattern? In other words, what is the correlation length of the random pattern  $I^{\text{lighting}}(z)$ ?

In this work, we make the following assumption:

**Approximation 3:** The correlation length of  $I^{\text{lighting}}(z)$  is much smaller than the attenuation length  $1/\eta$ .

This means that the caustic pattern changes from bright to dark features within short mutual distances along the LOS, while the attenuation weight does not change much in such mutual proximity. According to [30], a typical flicker correlation distance may vary between a few centimeters to a few tens of centimeters, depending on the underwater depth and the water surface waves. On the other hand, the attenuation distance typically varies [1] between a few meters to a few tens of meters. Therefore, there is typically at least 1 order of magnitude between the flicker correlation length and  $1/\eta$ .

Under approximation 3, the spatial variations of the caustic are filtered out by the integral in Eq. (10). Consequently, Eq. (5) approximately holds for the left camera, and, for the same reasons, also for the right camera. Hence,

$$\begin{aligned} B_R(\mathbf{x}_R, m) &\approx B_L(\mathbf{x}_L, m) \\ &= B_\infty(m) (1 - \exp\{-\eta[Z_{\text{obj}}(\mathbf{x}_L) - Z_h]\}) \quad \forall m. \end{aligned} \quad (11)$$

The temporal variations in  $B_\infty(m)$  are small, for the reason mentioned above: the flicker changes the distribution of energy, but not its spatial integral. Hence, in different frames, the pattern changes,

bright voxels dim and vice versa, but the integral over the LOS is usually insensitive to these temporal changes. Compounding Eqs. (6), (9), and (11), the overall scene radiance, as measured by the two stereo cameras can be formulated as

$$I_R(\mathbf{x}_R, m) \approx I_L(\mathbf{x}_L, m) \quad \forall m. \quad (12)$$

The limits and validity range of approximation 3 are analyzed using ray-tracing simulations described in Section 7.

## 4. Correspondence from Flicker

Equation (12) claims intensity similarity at points  $\mathbf{x}_R$  and  $\mathbf{x}_L$  at frame  $m$ . However, this similarity is generally not unique, at frame  $m$ . A set of pixels  $\Omega_R(m) = \{\mathbf{x}_R^{\text{incorrect}}\}$  in  $I_R$  have intensities that are very close to, or equal to  $I_L(\mathbf{x}_L)$ . One reason why this can happen is that objects at such noncorresponding pixels may have the same reflectance, irradiance and backscatter. This situation leads to the classic correspondence problem in nonflickering environments. More generally, the reflectance, irradiance and backscatter in  $\mathbf{x}_R^{\text{incorrect}}$  are all different than the ones in  $\mathbf{x}_L$ , but their combination in Eq. (6) yields the same overall intensity, at frame  $m$ .

Fortunately, in flicker, such ambiguities are completely resolved with high probability, since the lighting is dynamic. It can be claimed that flicker grants each object point a unique temporal signature that disambiguates stereo correspondence. Because of the lighting dynamics, noncorresponding pixels in  $\Omega_R(m)$  are generally different than those at  $\Omega_R(m')$ , in frame  $m' \neq m$ . A coincidence of matching intensities at  $m$  has rare chances of reoccurring at  $m'$ . Considering a large number of frames  $N_F$ ,

$$\bigcap_{m=1}^{N_F} \Omega_R(m) \rightarrow \emptyset, \quad (13)$$

where in practice even a small  $N_F$  suffices to eliminate the noncorresponding pixels.

Correspondence is determined in our work using simple temporal normalized correlation. Define the vector

$$\mathbf{I}_L(\mathbf{x}_L) \equiv [I_L(\mathbf{x}_L, 1), I_L(\mathbf{x}_L, 2), \dots, I_L(\mathbf{x}_L, N_F)]^T, \quad (14)$$

where T denotes transposition. Now, in the right image, there is a set of pixels  $\Psi$ , each of which is a candidate for correspondence with  $\mathbf{x}_L$ . Without calibration of the stereo setup,  $\Psi$  is the whole field of view (all the pixels in the right image). If calibration of the system had been done, then  $\Psi$  is the epipolar line [31,32] corresponding to  $\mathbf{x}_L$ . For a candidate pixel  $\mathbf{x}_R^{\text{cand}} \in \Psi$ , define

$$\begin{aligned} \mathbf{I}_R(\mathbf{x}_R^{\text{cand}}) &\equiv [I_R(\mathbf{x}_R^{\text{cand}}, 1), I_R(\mathbf{x}_R^{\text{cand}}, 2), \dots, I_R(\mathbf{x}_R^{\text{cand}}, N_F)]^T. \end{aligned} \quad (15)$$

Subtracting the mean of each vector, we obtain

$$\tilde{\mathbf{I}}_L(\mathbf{x}_L) = \mathbf{I}_L(\mathbf{x}_L) - \langle \mathbf{I}_L(\mathbf{x}_L) \rangle, \quad (16)$$

$$\tilde{\mathbf{I}}_R(\mathbf{x}_R^{\text{cand}}) = \mathbf{I}_R(\mathbf{x}_R^{\text{cand}}) - \langle \mathbf{I}_R(\mathbf{x}_R^{\text{cand}}) \rangle. \quad (17)$$

The empirical normalized correlation [33] between  $\mathbf{x}_L$  and  $\mathbf{x}_R^{\text{cand}}$  is

$$C(\mathbf{x}_R^{\text{cand}}) = [\tilde{\mathbf{I}}_L(\mathbf{x}_L)^T \cdot \tilde{\mathbf{I}}_R(\mathbf{x}_R^{\text{cand}})] / [\|\tilde{\mathbf{I}}_L(\mathbf{x}_L)\|_2 \|\tilde{\mathbf{I}}_R(\mathbf{x}_R^{\text{cand}})\|_2]. \quad (18)$$

For pixel  $\mathbf{x}_L$  in the left image, the corresponding pixel in the right image is then estimated as

$$\hat{\mathbf{x}}_R = \arg \max_{\mathbf{x}_R^{\text{cand}} \in \Psi} C(\mathbf{x}_R^{\text{cand}}). \quad (19)$$

As described in Section 3.A, the criterion of normalized temporal correlation is rather insensitive to intercamera differences in signal attenuation and object reflectance. It is also insensitive to differences in camera exposure parameters (gain and shutter speed) [34].

This method is very simple and accurate. It does not blur range edges, since it involves no spatial operations. However, it requires that correlation be established over a length of time. In static scenes, this is not a problem. However, it does not work well if camera and scene motions exist. In order to handle

a certain level of motion, the method can be enhanced by using spatiotemporal [21] rather than just temporal correlation. Furthermore, the acquisition time can be significantly reduced using regularization terms [35] to handle dynamic scenes, at the expense of pointwise accuracy.

## 5. Experiments

We conducted a set of *in situ* field experiments in the ocean and in a pool. Different scenes and cameras were used. In general,  $\Psi$  can be limited to a line in  $I_R$ , termed the epipolar line. Knowing the epipolar geometry requires calibration of the stereo rig, which is affected by a peculiar distortion [36]. Our work focuses on finding stereo correspondences, rather than calibration. Therefore, in all the following experiments, the setup is uncalibrated. Hence, the search domain  $\Psi$  includes the entire field of view.

### A. Swimming Pool Experiment

Consider our first example, which is an experiment conducted in a swimming pool. The scene includes several objects at  $Z_{\text{obj}} \in [1\text{ m}, 2\text{ m}]$ , near the corner of the pool. The depth at the bottom was  $\sim 1\text{ m}$ . The stereo setup was a Videre Design head shooting at 7 fps, with  $b = 25\text{ cm}$ . A sample frame-pair appears in Figs. 3(a) and 3(b). Temporal correlation was performed using  $N_F = 35$ . Examples of temporal matches in corresponding and noncorresponding points are shown in Figs. 3(c) and 3(d), respectively.

As common in studies dealing with stereo correspondence [23], the result is displayed as a disparity map, rather than a calibrated range map.

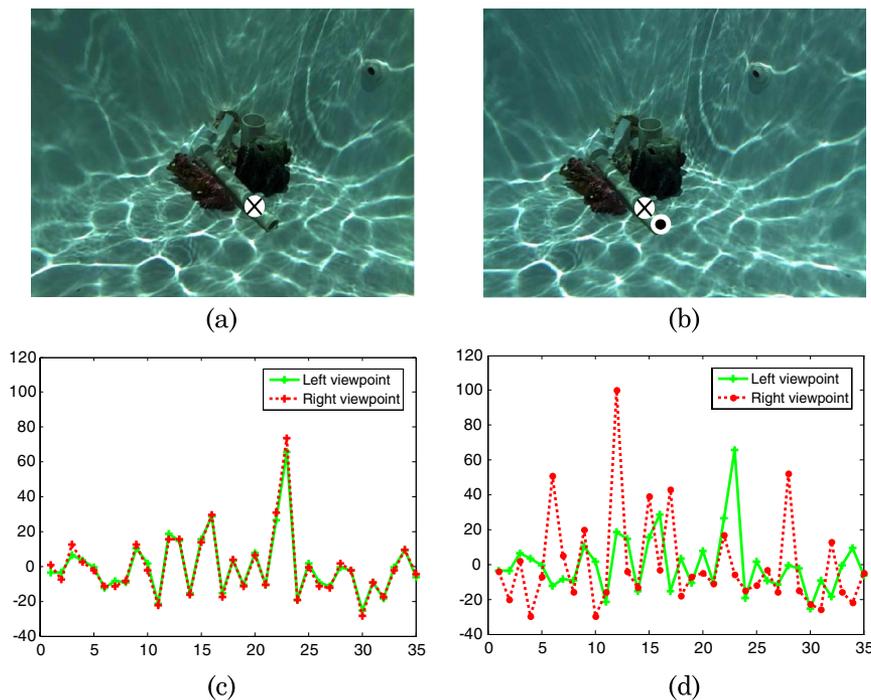


Fig. 3. (Color online) Left (a) and right (b) frames at one instance in the sequences. (c) Temporal plots of  $\tilde{\mathbf{I}}_L(\mathbf{x}_L)$  and  $\tilde{\mathbf{I}}_R(\hat{\mathbf{x}}_R)$  extracted from corresponding pixels. These pixels are marked by  $\otimes$  in the respective frames. (d) Temporal plots of  $\tilde{\mathbf{I}}_L(\mathbf{x}_L)$  and  $\tilde{\mathbf{I}}_R(\hat{\mathbf{x}}_R)$  extracted from noncorresponding pixels. These pixels are marked by  $\otimes$  and  $\circ$  in the respective frames.

The disparity map is derived based on Eq. (1):

$$\hat{d}(\mathbf{x}_L) = \|\hat{\mathbf{x}}_R - \mathbf{x}_L\|. \quad (20)$$

The inverse-disparity map derived in this experiment is displayed in Fig. 4.

### B. Oceanic Experiments

We conducted field experiments in the Mediterranean and the Red Sea, aided by scuba diving. The experiments were conducted at depths of 3–5 m. The stereo setup is shown in Fig. 1(b). Here, we used Canon HV-30 high-definition PAL video cameras in Ikelite underwater housings. To synchronize the video sequences, a blinking flashlight was shined into the running cameras before and after each experiment. These blinks were later detected in post-processing and used to temporally align the videos.

In the sea, the visibility was much poorer than in the pool. Hence, the flicker pattern had lower contrast. This required somewhat longer sequences to reliably establish the correlation, and thus correspondence. In any case, the sequences were just a few seconds long.

In one experiment, a natural scene in an underwater archeological site was captured using a  $b = 70$  cm baseline and  $N_F = 66$ . The resulting disparity map is presented in Fig. 5. The distance of the large cube from the cameras was  $\sim 5$  m. Some regions are marked in black. They represent low correspondence reliability, as we explain in Section 6.

Another experiment conducted in the Red Sea is shown in Fig. 6. Here, visibility was better than in the Mediterranean experiment. The baseline was

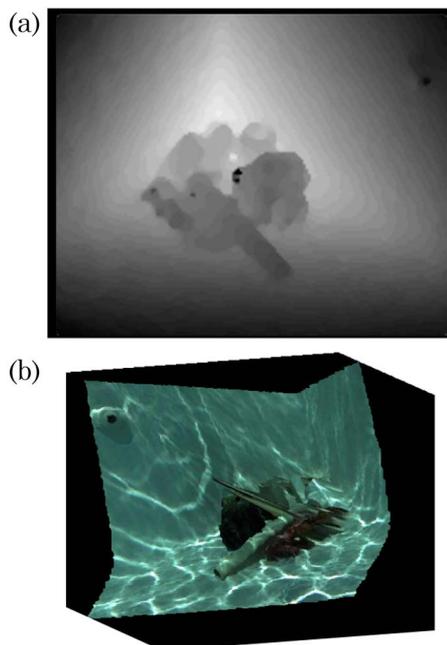


Fig. 4. (Color online) (a) The estimated inverse disparity map  $1/\hat{d}$  of the pool experiment. The result in (a) is used for texture mapping a different viewpoint in (b).

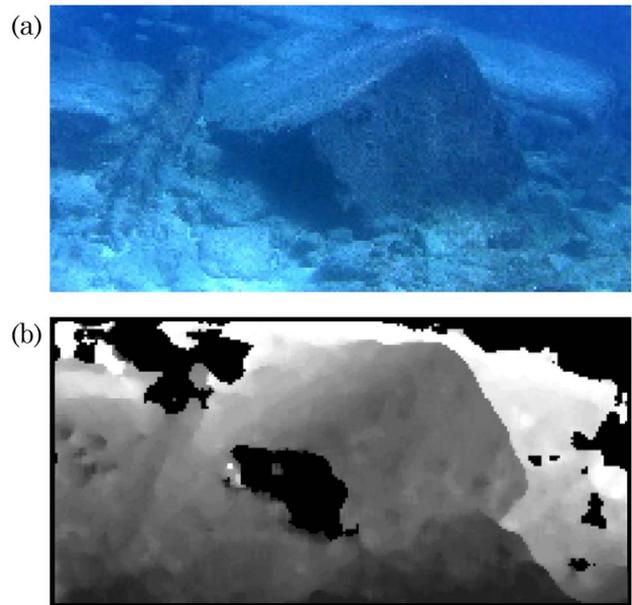


Fig. 5. (Color online) (a) Raw left frame from an experiment in a marine archaeological site (Caesarea). (b) The estimated inverse disparity map. Black areas represent low correspondence reliability, as explained in Section 6. Taken from [21] with permission ©2009 IEEE.

$b = 30$  cm and  $N_F = 50$ . The distances of the bowl, the board and the chair were 2 m, 2.5 m and 3 m, respectively.

### 6. Limitations Because of Shadows and Occlusions

It is possible to assess at which image locations correspondence estimation using Eq. (18) is unreliable. As in any stereo setup, occluded regions cannot be directly matched. More particular to this approach, however, is that some pixels simply correspond to object points that reside in the shadow of downwelling lighting, due to objects above them. Points in the shadow are unaffected by flicker and thus

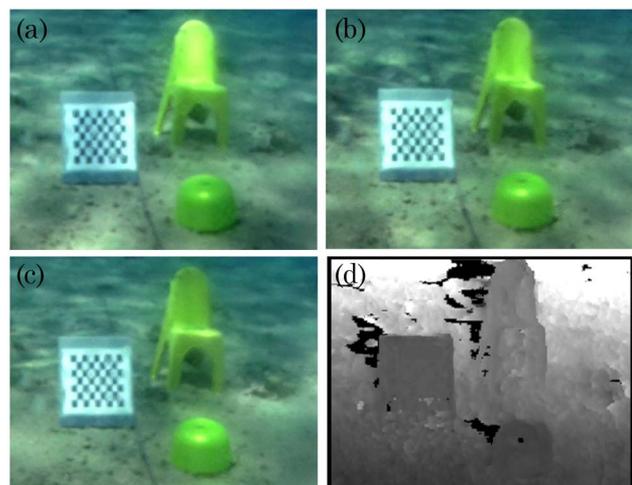


Fig. 6. (Color online) (a)–(c) Raw left frames from the Red Sea experiment. (d) The estimated inverse disparity map. Black areas represent low correspondence reliability, as explained in Section 6.

contain no temporal modulation. Therefore, the resulting correspondence estimation in shadowed regions is unreliable, if based solely on temporal correlation.

The set of pixels in  $\mathbf{I}_L$  that are occluded in  $\mathbf{I}_R$  have a low value of  $C$  even in the optimal match  $\hat{\mathbf{x}}_R$ . Hence, if  $C(\hat{\mathbf{x}}_R)$  is below a threshold  $\chi_C$ , it indicates the correspondence is unreliable. In pixels in which flicker is absent or too weak to detect, the standard deviation (STD) of the pixel value  $\|\tilde{\mathbf{I}}_L(\mathbf{x}_L)\|_2$  is very low. Hence, this set can be assessed by thresholding the field  $\|\tilde{\mathbf{I}}_L(\mathbf{x}_L)\|_2$  by a parameter  $\chi_{\text{STD}}$ .

Thus, we may define a set  $\Psi_{\text{reliable}}$  of reliable pixels.

$$\Psi_{\text{reliable}} = \{\mathbf{x}_L: [C_{\mathbf{x}_L} > \chi_C] \text{ AND } \|\tilde{\mathbf{I}}_L(\mathbf{x}_L)\|_2 > \chi_{\text{STD}}\}. \quad (21)$$

Range information in pixels outside  $\Psi_{\text{reliable}}$  should be filled-in using other mechanisms, such as regularization [35] or inpainting.

## 7. Limitations Because of Backscatter

### A. Motivation

As explained in Section 3.B, backscatter under flickering illumination varies in time, space and viewpoint. Let us consider a pair of corresponding points in two viewpoints. The temporal signal of these imaged points should have a strong match and achieve a high correlation score. However, the backscatter pattern varies between the two viewpoints. Therefore, backscatter variations are considered a random disturbance to the correlation of corresponding pairs. This may affect the matching score between corresponding object points.

In Section 3.B, approximation 3 neglects backscatter variations. It assumes that these variations are filtered out by the LOS integral. In the following sections, a simulation is used to test this approximation and indicate limitations of approximation 3 and thus of the correspondence-finding principle. The simulation focuses on coastal ocean sea surface and water volume properties. These properties may be different in other scenarios, such as pools and small lakes. Thus, to better study the limitations in such places, the simulated model (specifically, the surface statistics) may need to be adapted.

### B. Statistics

In a simulated scene, the signal and backscatter images are  $S(\mathbf{x}, m)$  and  $B(\mathbf{x}, m)$ , respectively. The temporally averaged signal  $\bar{S}(\mathbf{x})$  is

$$\bar{S}(\mathbf{x}) = (1/N_F) \sum_{m=1}^{N_F} S(\mathbf{x}, m). \quad (22)$$

The STD of the signal is estimated by

$$\sigma_S(\mathbf{x}) = \left\{ (1/N_F) \sum_{m=1}^{N_F} [S(\mathbf{x}, m) - \bar{S}(\mathbf{x})]^2 \right\}^{1/2}. \quad (23)$$

The definitions of  $\bar{B}(\mathbf{x})$  and  $\sigma_B(\mathbf{x})$  are analogously set by substituting  $B$  instead of  $S$  in Eqs. (22) and (23). We spatially average these quantities, per constant object range  $\tilde{Z}_{\text{obj}}$ . For example,

$$\sigma_S(\tilde{Z}_{\text{obj}}) = \frac{1}{|\Theta(\tilde{Z}_{\text{obj}})|} \sum_{\mathbf{x} \in \Theta(\tilde{Z}_{\text{obj}})} \sigma_S(\mathbf{x}), \quad (24)$$

where

$$\Theta(\tilde{Z}_{\text{obj}}) \equiv \{\mathbf{x}: Z_{\text{obj}}(\mathbf{x}) = \tilde{Z}_{\text{obj}}\}. \quad (25)$$

Similarly,  $\bar{B}(\tilde{Z}_{\text{obj}})$ ,  $\bar{S}(\tilde{Z}_{\text{obj}})$  and  $\sigma_B(\tilde{Z}_{\text{obj}})$  are derived.

Now, consider the ratio

$$\rho_S(\tilde{Z}_{\text{obj}}) = \sigma_S(\tilde{Z}_{\text{obj}}) / \bar{S}(\tilde{Z}_{\text{obj}}). \quad (26)$$

If one simply desires to capture a flicker-free image, then  $\bar{S}$  represents the signal while the STD  $\sigma_S(\tilde{Z}_{\text{obj}})$  indicates the noise in the signal. Hence, for flicker-free imaging, the signal-to-noise ratio (SNR) is  $1/\rho_S$ . However, our task is not a flicker-free image, but establishing stereo correspondence. In this task,  $\sigma_S$  is not noise: it is the key for achieving the task. The higher  $\sigma_S$  is, the better. So, for our task,  $\sigma_S$  is the flicker signal. What is flicker noise? In our context, noise is random radiance that counters approximation 3. In other words,  $\sigma_B$  is the flicker noise. To assess the SNR in the context of our task, we consequently define a flicker signal-to-noise ratio (FSNR) as

$$\text{FSNR}(\tilde{Z}_{\text{obj}}) = \sigma_S(\tilde{Z}_{\text{obj}}) / \sigma_B(\tilde{Z}_{\text{obj}}). \quad (27)$$

We use the FSNR to assess the range in which the flicker-signal  $\sigma_S$  is dominant, relative to the flicker-noise  $\sigma_B$ . In this range, we expect to achieve good correspondence using temporal correlation. In the next sections, the simulation steps that yield these statistics are described.

### C. Simulation Description

First, we describe the principles used in the ray-tracing simulation. Specific dimensions and parameters are listed in Section 7.E. The simulation is divided into four parts:

1. Setting parameters: Scene (object) structure, medium properties, illumination properties and data structures.
2. Dynamic data: Water surface generation and dynamics [37,38].
3. Ray-tracing simulation including reflection, refraction, attenuation, and scattering.
4. Viewpoint generation: Signal and backscatter calculations, per viewpoint.

We now elaborate on these parts.

### 1. Setting Parameters

The scene is set by some parameters. These include the dimensions of the water tank, voxel resolution and initial density of traced rays. In addition, a water surface is generated. It is possible to measure the water surface, for example, using sun or laser glitter [39,40]. However, in this work we do not measure the surface, but simulate it using a sea surface model and statistics, as described in [37]. Further details regarding the surface simulation appear in Appendix A.

### 2. Illumination Ray-tracing

In this part, light rays are generated and propagated. Figure 7 illustrates the processes of surface generation and ray tracing.

A ray is indexed by  $w$ . Let  $\hat{c}(w, t)$  be the direction unit vector of ray  $w$  at time  $t$ . The speed of light in the medium is  $c_n$ , where  $n$  denotes the optical refractive index of the medium (air or water). Ray propagation is divided into small temporal steps  $\Delta t$ . At time  $t$ , the position of propagating ray  $w$  is  $\mathbf{y}(w, t)$ . Then,

$$\mathbf{y}(w, t + \Delta t) = \mathbf{y}(w, t) + c_n \hat{c}(w, t) \Delta t. \quad (28)$$

All rays are initialized outside the water, carrying equal irradiance. When ray  $w$  encounters the water surface, the direction vector  $\hat{c}$  changes and its associated radiance is attenuated according to Snell's and Fresnel's laws, respectively. In the water, the ray radiance is attenuated according to the attenuation coefficient and the length of each step:

$$I^{\text{lighting}}(w, t + \Delta t) = I^{\text{lighting}}(w, t) \exp(-\eta c_n \Delta t). \quad (29)$$

Ray tracing continues until all the rays either exit the water tank or get absorbed at the bottom.

The simulated space is divided into small voxels, each indexed by  $v$ . We set the step  $c_n \Delta t$  to be much smaller than the voxel length (see Fig. 8). This generally creates several recorded instances of ray  $w$  in any voxel that  $w$  passes through. This reduces numerical errors. Define the set of rays

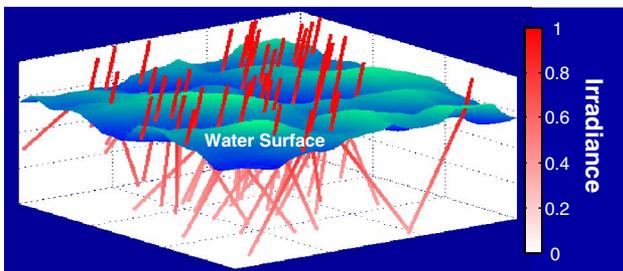


Fig. 7. (Color online) A simulated water surface having random light rays refracted and attenuated. The faded red color represents attenuation of normalized irradiance, due to absorption and scattering.

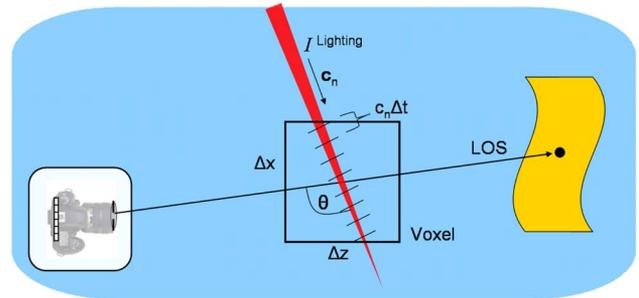


Fig. 8. (Color online) A light ray passing through a voxel. A record of irradiance and direction of the ray at each temporal step is kept. The angle between LOS and the ray  $\theta$  is calculated for backscatter calculations.

$$\Omega_v(t) \equiv \{w: \mathbf{y}(w, t) \in v\}. \quad (30)$$

A data structure lists all the values of  $I^{\text{lighting}}(w, t)$  that correspond to voxel  $v$

$$\Upsilon_v^{\text{lighting}} = \bigcup_t \bigcup_{w \in \Omega_v(t)} I^{\text{lighting}}(w, t). \quad (31)$$

An additional data structure lists all the corresponding vectors  $\hat{c}(w, t)$  in  $v$ :

$$\Upsilon_v^{\text{directions}} = \bigcup_t \bigcup_{w \in \Omega_v(t)} \hat{c}(w, t), \quad (32)$$

These data structures enable consequent calculation of the signal and backscatter images in any viewpoint.

### 3. Backscatter and Signal Calculations

This part deals with scattering in each voxel, following data accumulated in Eqs. (31) and (32). Camera properties such as field of view, resolution, location and orientation are accounted for. For each pixel in each camera, the backscatter integral (10) is discretized to a sum over all the voxels on the LOS corresponding to that pixel and camera. This results in backscatter images. Signal images are calculated separately using irradiance in voxels that reside at the bottom of the water tank. The signal calculations assume the surfaces to be Lambertian, with a certain albedo.

Finally, the signal and backscatter images are added, thus creating a rendered underwater image. The attenuation and scattering coefficients vary as function of light wavelength [1,41]. Thus, we repeat the entire ray-tracing process three times, each using different coefficients, corresponding to red, green and blue light wavelengths. The result of this process is a color image of the underwater scene, as shown in Fig. 9.

#### D. Sky Radiance

Without direct sunlight, underwater flicker is very weak. Skylight contributes to underwater ambient lighting, but has very little effect on underwater

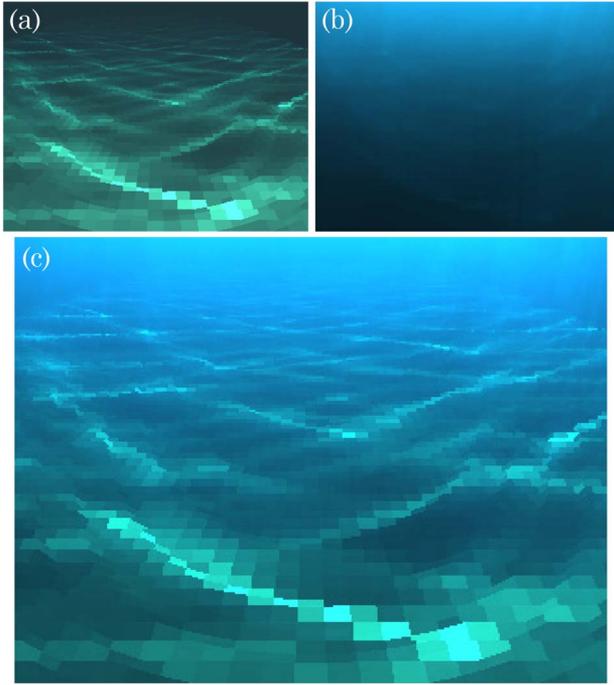


Fig. 9. (Color online) (a) Rendered signal. (b) Rendered backscatter pattern. (c) Rendered color image of an underwater scene.

flicker. In order to demonstrate this, we simulated different types of overcast International Commission on Illumination standard sky radiance [42], refracted through the wavy water surface, as described in Section 7.C.2. The relative skylight and sunlight variations were compared using different physical parameters. Overall, considering Eq. (26), our simulation yielded  $\rho_S^{\text{Sun}} \in [60\%, 80\%]$ , while  $\rho_S^{\text{Sky}} \in [0.5\%, 2\%]$ . As for backscatter variations (Eq. (27)),  $\rho_B^{\text{Sun}} \in [2\%, 10\%]$ , while  $\rho_B^{\text{Sky}} \in [0.06\%, 0.2\%]$ . We are interested in signal and backscatter variations. Therefore, ambient skylight and its contribution to flicker variations can be neglected.

#### E. Simulation and Results

The following figures present results of a simulation using attenuation coefficients of coastal ocean [41,43]:  $(\eta_{450\text{nm}}, \eta_{550\text{nm}}, \eta_{650\text{nm}}) = (0.22, 0.23, 0.49)$ , respectively. The zenith angle of the sun is set to be  $20^\circ$  and the azimuth angle with respect of the LOS is  $90^\circ$ . The water depth is 3 m and wind speed is 4 m/Sec [38]. The voxel resolution is 0.1 m. The ray-tracing resolution (step) is 0.01 m. The bottom of the sea is set to have spectral reflectance of a quartz beach sand [44]. The phase function  $p(\theta)$  we used [41,45], is

$$p(\theta) = (1 - \varphi^2) / [4\pi(1 + \varphi^2 - 2\varphi \cos \theta)^{1.5}], \quad (33)$$

where  $\varphi = 0.924$ , following [41].

The following results are based on statistics of 600 images of the fixed scene. In Fig. 10,  $\sigma_S$ ,  $\bar{S}$  and  $\rho_S$  of the sea bottom are plotted as a function of  $Z_{\text{obj}}$ , for three light wavelengths.

It can be noticed that there is an exponential decay of  $\bar{S}$  as function of  $Z_{\text{obj}}$ . This decay is consistent with the image formation model [Eq. (3)]. Moreover, Fig. 10(c) shows that approximately  $\sigma_S \propto \bar{S}$ . This means that  $\sigma_S$  can be used as a good indicator for  $\bar{S}$ .

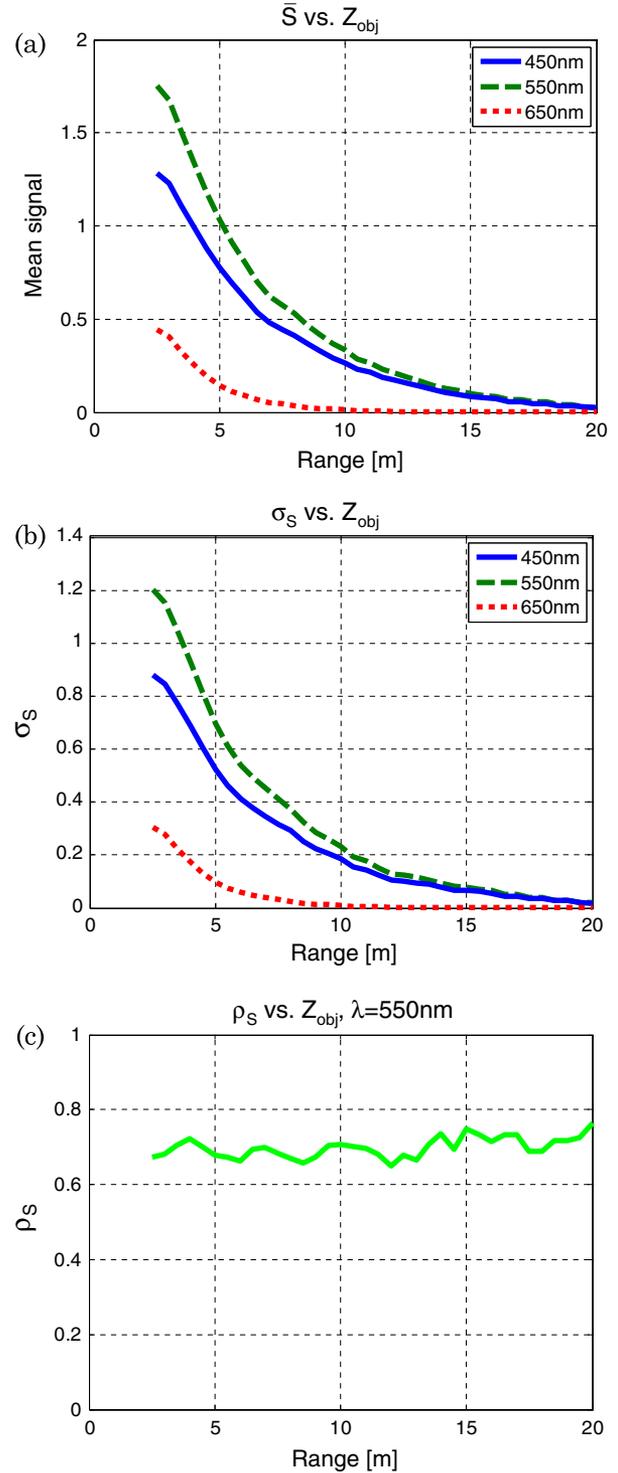


Fig. 10. (Color online) (a)  $\bar{S}$  versus  $Z_{\text{obj}}$ , for different light wavelengths. (b)  $\sigma_S$  versus  $Z_{\text{obj}}$ , for different light wavelengths. (c)  $\rho_S$  versus  $Z_{\text{obj}}$ ,  $\lambda = 550\text{nm}$ .

Figure 11 presents  $\sigma_B$ ,  $\bar{B}$  and  $\rho_B$  of the same scene. Here there is exponential saturation of  $\bar{B}$  as a function of  $Z_{obj}$ , which is also consistent with the image formation model [Eq. (5)]. Moreover,  $\sigma_B$  is almost constant as a function of  $Z_{obj}$ . This is due to the fact that the majority of backscatter variations are contributed by nearby parts of the LOS.

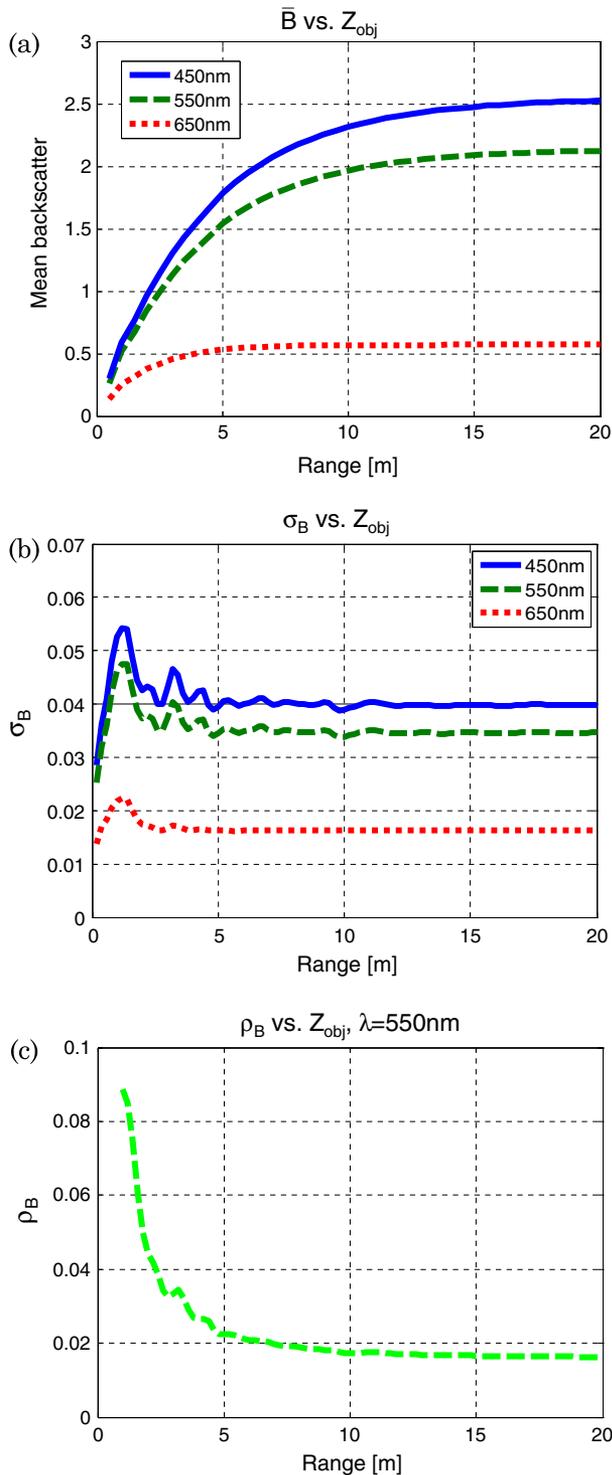


Fig. 11. (Color online) (a)  $\bar{B}$  versus  $Z_{obj}$ , of different light wavelengths. (b)  $\sigma_B$  versus  $Z_{obj}$ , of different light wavelengths. (c)  $\rho_B$  versus  $Z_{obj}$ ,  $\lambda = 550$  nm.

Another interesting matter is revealed by  $\rho_S$  and  $\rho_B$ . In Figs. 10(c) and 11(c),  $\rho_S \in [60\%, 70\%]$ , while  $\rho_B \in [2\%, 10\%]$  only. This raises an important observation:  $\bar{S}$  is often of the same order of magnitude as  $\bar{B}$ , but the flicker signal from objects is much more prominent than flicker noise, since  $\sigma_S \gg \sigma_B$ . This observation supports our approximation (Section 3.B) and emphasizes the advantage of using temporal image variation, instead of the image itself, for stereo. Figure 12(a) presents  $\text{FSNR}(Z_{obj})$ .

#### F. FSNR and Performance Assessment

Which FSNR level is required to assure good correspondence? To assess this, another viewpoint was rendered, to simulate stereoscopic vision. Correspondence between simulated viewpoints was found using temporal correlation and compared to the known ground truth. Figure 12(b) presents the rate of successful correspondence as function of FSNR, for various values of  $N_F$ . As a rule of thumb, more than 90% correct correspondence is achieved when  $\text{FSNR} > \text{FSNR}_T$ , where  $\text{FSNR}_T \cong 5$ . This is based

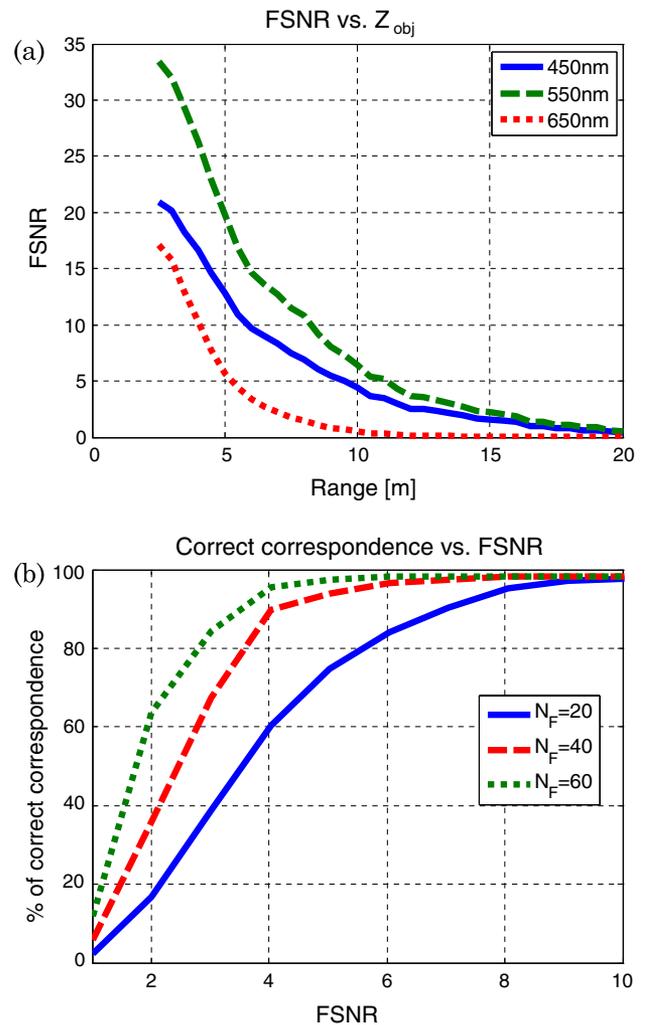


Fig. 12. (Color online) (a) FSNR versus  $Z_{obj}$ , of different light wavelengths. (b) The rate of successful match versus FSNR,  $\lambda = 550$  nm.

solely on temporal variations of flicker, and is independent of the spatial reflectivity texture. If the objects are textured with high contrast and spatial frequency, the high rate of correct correspondence is achieved using even a lower FSNR.

Observing Figs. 12(a) and 12(b), a good FSNR (yielding a high rate of good correspondence) can be observed up to  $Z_{\text{obj}}$  between  $2\eta^{-1}$  and  $3\eta^{-1}$ . This observation may be used as a practical rule of thumb for a lower bound for the validity range of the method.

## 8. Discussion

The method presented in this paper exploits the natural phenomenon of underwater flicker to create a dense correspondence map. This is related to [46,47], where controlled artificial lighting patterns used a projector to assist stereo video cameras. The effectiveness of our method is demonstrated by *in situ* experiments in the field. In order to explore the theoretical limits of the method, a ray-tracing simulation was implemented. The method establishes correspondence rather reliably even without epipolar constraints. So, in turn, the results of a sequence of such correspondence mappings can possibly establish the epipolar geometry of the system. The method strongly depends on direct sunlight. Therefore, correspondence results in shadowed areas are unreliable. In order to overcome this problem, spatial support [21] and regularization [35] should be considered.

The study of stereo vision under flicker may be relevant to biological vision. Consider marine animals that share three properties: (1) They live in shallow water, where there is abundance of natural light and thus flicker; (2) They have binocular vision, which can potentially enable them to assess distance by triangulating objects from different eye elements; (3) Their brain is very small. Can such small brains solve the complex problems associated with correspondence? Specifically, such animals include stomatopods and other crustaceans (e.g., lobsters) [8,48]. The mechanism proposed in this paper may potentially assist stereo vision in such animals: it suits their habitat, and requires very simple, point-wise calculations. We have no supporting evidence for such a hypothesis, but it is worth exploring. For instance, the animals may undergo controlled lab tests, in a task that requires distance assessment (prey, navigation). Their performance under flicker (wavy water surface, or artificially projected illumination) can be compared to the success in still water and lighting. Interestingly, the spatial and temporal frequencies of flicker match the contrast sensitivity and temporal visual response [49] of some marine animals.

## Appendix A—Sea Surface Model

We use a surface model described in [37]. To make the paper self-contained, we describe the steps of sea surface simulation. Let  $\mathbf{r}$  be a horizontal coordinate in the world. Then, define  $h(\mathbf{r})$  to be the

surface height relative to a flat sea surface at  $\mathbf{r}$ . The autocorrelation function of  $h(\mathbf{r})$  is

$$\xi(\Delta\mathbf{r}) = \langle h(\mathbf{r}), h(\mathbf{r} + \Delta\mathbf{r}) \rangle, \quad (\text{A1})$$

where  $\Delta\mathbf{r}$  is a horizontal displacement vector on the surface. The roughness spectrum of the surface is the spatial Fourier transform of  $\xi(\Delta\mathbf{r})$ :

$$\Gamma(\mathbf{k}) = \mathcal{F}\{\xi(\Delta\mathbf{r})\}. \quad (\text{A2})$$

Here  $\mathbf{k} = (k, \phi)$  in polar coordinates, where  $k$  is the surface wave number, and  $\phi$  is the propagation direction of the surface wave mode.

The spatial spectrum is a product of an omnidirectional spectral distribution and an angular spread

$$\Gamma(k, \phi) = k^{-1}\nu(k)\Phi(k, \phi), \quad (\text{A3})$$

as described in [37,38]. The functions  $\nu(k)$  and angular spreading  $\Phi(k, \phi)$  depend on wind speed, direction and water depth [38,50]. Surface waves propagate in a dispersive manner. The temporal frequency  $\omega_k$  of a surface wave with wave-number  $k$  is given [37] by

$$\omega_k = \sqrt{gk \tanh(kD) + T_W \zeta_W^{-1} k^3}. \quad (\text{A4})$$

Here  $D$  is the water depth, while  $T_W = 0.0728$  [N/m] and  $\zeta_W = 997$  [kg/m<sup>3</sup>] are the water surface tension and density, respectively, and  $g$  is the gravitational acceleration.

We would like to produce a random amplitude, based on the roughness spectrum. Therefore, for each  $\mathbf{k}$ , a random amplitude  $\mu_{\mathbf{k}}$  and phase  $\kappa_{\mathbf{k}}$  are drawn. These random variables are uniformly distributed in  $[0, 1]$  and  $[0, 2\pi]$ , respectively. The sea surface at each moment  $\tau$  is defined in the Fourier plane as a dispersive sum of waves propagating in opposite directions [37],

$$\tilde{H}(\mathbf{k}, \tau) = \tilde{H}_0(\mathbf{k}) \exp(i\omega_k \tau) + \tilde{H}_0^*(-\mathbf{k}) \exp(-i\omega_k \tau), \quad (\text{A5})$$

where

$$\tilde{H}_0(\mathbf{k}) = \sqrt{2}^{-1} \mu_{\mathbf{k}} \exp(i\kappa_{\mathbf{k}}) \sqrt{\Gamma(\mathbf{k})}. \quad (\text{A6})$$

Finally, the sea surface height  $h(\mathbf{r}, \tau)$  at each moment is calculated by an inverse Fourier transform of  $\tilde{H}(\mathbf{k}, \tau)$

$$h(\mathbf{r}, \tau) = \mathcal{F}^{-1}\{\tilde{H}(\mathbf{k}, \tau)\}. \quad (\text{A7})$$

Once random initialization is done in Eq. (A6), the surface wave propagates deterministically.

We thank Tali Treibitz, Izchak Yogev, Amin Sarafraz and Ali Taatian for their help in conducting the underwater experiments and Ayellet Tal for useful discussions. We thank Nadav Shashar, Justin

Marshall, Tom Cronin and Dan Nilsson for helpful discussions about potential stereo vision in marine invertebrates. Yoav Schechner is a Landau Fellow—supported by the Taub Foundation. The work was supported by the Israel Science Foundation (Grant 1031/08) and the US-Israel Binational Science Foundation (Grant 2006384). This work was conducted in the Ollendorff Minerva Center. Minerva is funded through the BMBF.

## References

1. R. E. Walker, *Marine Light Field Statistics* (John Wiley, 1994), Chap. 10.
2. N. G. Jerlov, *Marine Optics* (Elsevier, 1976), Chap. 6.
3. C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (John Wiley, 1983).
4. A. Tonizzo, J. Zhou, A. Gilerson, M. S. Twardowski, D. J. Gray, R. A. Arnone, B. M. Gross, F. Moshary, and S. A. Ahmed, "Polarized light in coastal waters: hyperspectral and multianalytical analysis," *Opt. Express* **17**, 5666–5683 (2009).
5. N. Gracias, S. Negahdaripour, L. Neumann, R. Prados, and R. Garcia, "A motion compensated filtering approach to remove sunlight flicker in shallow water images," in *Proc. MTS/IEEE Oceans* (IEEE, 2008).
6. D. K. Lynch and W. Livingston, *Color and Light in Nature*, 2nd ed. (Cambridge University, 2001), Chaps. 3,4.
7. H. R. Gordon, "Normalized water-leaving radiance: revisiting the influence of surface roughness," *Appl. Opt.* **44**, 241–248 (2005).
8. T. W. Cronin and N. J. Marshall, "Parallel processing and image analysis in the eyes of mantis shrimps," *Biol. Bull.* **200**, 177–183 (2001).
9. S. Johnsen, N. J. Marshall, and E. A. Widder, "Polarization sensitivity as a contrast enhancer in pelagic predators: lessons from in situ polarization imaging of transparent zooplankton," *Phil. Trans. R. Soc. B* **366**, 655–670 (2011).
10. D. C. Parkyn, J. D. Austin, and C. W. Hawryshyn, "Acquisition of polarized-light orientation in salmonids under laboratory conditions," *Anim. Behav.* **65**, 893–904 (2003).
11. G. Horvath and D. Varju, "On the structure of the aerial visual field of aquatic animals distorted by refraction," *Bull. Math. Biol.* **53**, 425–441 (1991).
12. L. M. Mathger, N. Shashar, and R. T. Hanlon, "Do cephalopods communicate using polarized light reflections from their skin?" *J. Exp. Biol.* **212**, 2133–2140 (2009).
13. A. Sarafraz, S. Negahdaripour, and Y. Y. Schechner, "Enhancing images in scattering media utilizing stereovision and polarization," in *Proceedings of the IEEE Workshop Applications of Computer Vision* (IEEE, 2009).
14. D. M. Kocak, F. R. Dalgleish, F. M. Caimi, and Y. Y. Schechner, "A focus on recent developments and trends in underwater imaging," *Mar. Technol. Soc. J.* **42**, 52–67 (2008).
15. M. Bryant, D. Wettergreen, S. Abdallah, and A. Zelinsky, "Robust camera calibration for an autonomous underwater vehicle," in *Proceedings of Australian Conference on Robotics and Automation* (2000), pp. 111–116.
16. J. M. Lavest, F. Guichard, and C. Rousseau, "Multiview reconstruction combining underwater and air sensors," in *Proceedings of the IEEE International Conference on Image Processing* (IEEE, 2000), pp. 813–816.
17. Y. Kahanov and J. Royal, "Analysis of hull remains of the Dor D vessel, Tantura Lagoon, Israel," *Int. J. Naut. Arch.* **30**, 257–265 (2001).
18. Y. Y. Schechner and N. Karpel, "Recovery of underwater visibility and structure by polarization analysis," *IEEE J. Ocean. Eng.* **30**, 570–587 (2005).
19. T. Treibitz and Y. Y. Schechner, "Active polarization descattering," *IEEE Trans. Pattern Anal. Machine Intell.* **31**, 385–399 (2009).
20. Y. Y. Schechner and N. Karpel, "Attenuating natural flicker patterns," in *Proceedings of MTS/IEEE Oceans* (IEEE, 2004), pp. 813–816.
21. Y. Swirski, Y. Y. Schechner, B. Herzberg, and S. Negahdaripour, "Stereo from flickering caustics," in *Proceedings of the IEEE International Conference on Computer Vision* (IEEE, 2009), pp. 205–212.
22. Y. Swirski, Y. Y. Schechner, B. Herzberg, and S. Negahdaripour, "Underwater stereo using natural flickering illumination," in *Proceedings of MTS/IEEE Oceans* (IEEE, 2010).
23. D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.* **47**, 7–42 (2002).
24. A. Fournier and W. T. Reeves, "A simple model of ocean waves," in *Proceedings of Special Interest Group on Computer Graphics and Interactive Technology*, pp. 75–84 (1986).
25. M. N. Gamito and F. K. Musgrave, "An accurate model of wave refraction over shallow water," *Comput. Graphics* **26**, 291–307 (2002).
26. B. Jahne, J. Klinke, and S. Waas, "Imaging of short ocean wind waves: a critical theoretical review," *J. Opt. Soc. Am.* **11**, 2197–2209 (1994).
27. J. S. Jaffe, "Computer modelling and the design of optimal underwater imaging systems," *IEEE J. Ocean. Eng.* **15**, 101–111 (1990).
28. J. E. Hansen and L. D. Travis, "Light scattering in planetary atmospheres," *Space Sci. Rev.* **16**, 527–610 (1974).
29. H. Zhang and K. J. Voss, "Bidirectional reflectance study on dry, wet, and submerged particulate layers: effects of pore liquid refractive index and translucent particle concentrations," *Appl. Opt.* **45**, 8753–8763 (2006).
30. A. Fraser, R. Walker, and F. Jurgens, "Spatial and temporal correlation of underwater sunlight fluctuations in the sea," *IEEE J. Ocean. Eng.* **5**, 195–198 (1980).
31. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision* (Cambridge University, 2003), Chaps. 9–12.
32. E. Trucco and A. Verri, *Introductory Techniques For 3D Computer Vision* (Prentice Hall, 1998), Chap. 6.
33. R. Eustice, O. Pizarro, H. Singh, and J. Howland, "UWIT: underwater image toolbox for optical image processing and mosaicking in Matlab," in *Proceedings of International Symposium on Underwater Technology* (IEEE, 2002), pp. 141–145.
34. A. Troccoli, S. B. Kang, and S. M. Seitz, "Multi-view multi-exposure stereo," in *Proceedings of 3D Data Processing, Visualization, and Transmission* (2006), pp. 861–868.
35. Y. Swirski, Y. Y. Schechner, and T. Nir, "Variational stereo in dynamic illumination," to appear in *Proceedings of the IEEE International Conference on Computer Vision* (IEEE, 2011).
36. T. Treibitz, Y. Y. Schechner, and H. Singh, "Flat refractive geometry," in *Proceedings of IEEE Computer Vision and Pattern Recognition* (IEEE, 2008).
37. J. Miranda, A. Camps, J. Gomez, M. Vall-llossera, and R. Villarino, "Time-dependent sea surface numerical generation for remote sensing applications," in *Proceedings of International Geographic and Remote Sensing Symposium* (IEEE, 2005), pp. 2527–2530.
38. S. Durden and J. Vesecky, "A physical radar cross-section model for a wind-driven sea with swell," *IEEE J. Ocean. Eng.* **10**, 445–451 (1985).
39. C. Cox and W. Munk, "Measurement of the roughness of the sea surface from photographs of the sun's glitter," *J. Opt. Soc. Am.* **44**, 838–850 (1954).

40. J. A. Shaw and J. H. Churnside, "Scanning-laser glint measurements of sea-surface slope statistics," *Appl. Opt.* **36**, 4202–4213 (1997).
41. C. D. Mobley, *Light and Water: Radiative Transfer in Natural Waters* (Academic, 1994), Chaps. 3–5.
42. S. Darula and R. Kittler, "CIE general sky standard defining luminance distributions," in *Proceedings of the International Building Performance Simulation Association* (IEEE, 2002).
43. R. W. Preisendorfer, *Hydrologic Optics* (U. S. Department of Commerce, 1976), Chaps. 1 and 12.
44. D. E. Bowker, R. E. Davis, D. L. Myrick, K. Stacy and W. T. Jones, "Spectral reflectances of natural targets for use in remote sensing studies," NASA Reference Publication A 1139 (1985), p. 141.
45. K. Kamiuto, "Study of the Henyey-Greenstein approximation to scattering phase functions," *J. Quant. Spectrosc. Radiat. Transfer* **37**, 411–413 (1987).
46. J. Davis, D. Nehab, R. Ramamoorthi, and S. Rusinkiewicz, "Spacetime stereo: a unifying framework for septh from triangulation," *IEEE Trans. Pattern Anal. Machine Intell.* **27**, 296–302 (2005).
47. L. Zhang, B. Curless, and S. M. Seitz, "Spacetime stereo: shape recovery for dynamic scenes," in *Proceedings IEEE Computer Vision and Pattern Recognition* (IEEE, 2003).
48. T. W. Cronin, J. N. Nair, R. D. Doyle, and R. L. Caldwell, "Ocular tracking of rapidly moving visual targets by stomatopod crustaceans," *J. Exp. Biol.* **138**, 155–179 (1988).
49. W. N. McFarland and E. R. Loew, "Wave produced changes in underwater light and their relations to vision," *Envir. Biol. Fish* **8**, 173–184 (1983).
50. A. Fung and K. Lee, "A semi-empirical sea-spectrum model for scattering coefficient estimation," *IEEE J. Ocean. Eng.* **7**, 166–176 (1982).