Underwater Stereo Using Natural Flickering Illumination

Yohay Swirski, Yoav Y. Schechner, Ben Herzberg
Dept. of Electrical Engineering
Technion - Israel Inst. Technology
Haifa 32000, Israel
yohays@tx.technion.ac.il,
yoav@ee.technion.ac.il,
bherzberg@gmail.com

Shahriar Negahdaripour
Electrical and Computer Eng. Dept.
University of Miami
Coral Gables, FL 33124-0640
nshahriar@miami.edu

***Abstract*—Computer vision is challenged by the underwater environment. Poor visibility, geometrical distortions and non-uniform illumination typically make underwater vision less trivial than open air vision. One effect which can be rather strong in this domain is *sunlight flicker*. Here, submerged objects and the water volume itself are illuminated in a natural random pattern, which is spatially and temporally varying. This phenomenon has been considered mainly as a significant *disturbance* to vision. We show that the spatiotemporal variations of flicker can actually be *beneficial* to underwater vision. Specifically, flicker disambiguates stereo correspondence. This disambiguation is very simple, yet it yields accurate results. Under flickering illumination, each object point in the scene has a unique, unambiguous temporal signature. This temporal signature enables us to find dense and accurate correspondence underwater. This process may be enhanced by involving the spatial variability of the flicker field in the solution. The method is demonstrated underwater by in-situ experiments. This method may be useful to a wide range of shallow underwater applications.**

I. Introduction

Computer vision is applied underwater [1], [2], [3] in a wide range of tasks, including robotic operations [4], [5] and inspection of cables and pipelines [6]. Moreover, optical underwater imaging is also applied to archaeological documentation [7] and observation of wildlife [8], [9], [10]. There is a significant role for computer vision in shallow water [11], e.g., for inspection of ports, ship hulls [12] and monitoring swimming pools [13].

One effect which can be rather strong in this domain is *sunlight flicker*. Here, submerged objects and the water volume itself are illuminated by a natural random pattern [14], [15] which is spatially and temporally varying. An example is shown in Fig. 1. This phenomenon has mostly been considered as a significant visual *disturbance*. Thus, attempts were made to reduce this effect by postprocessing [16], [17].

In Ref. [18], it was noted that the spatiotemporal variations of flicker can actually be very beneficial to underwater vision. Specifically, flicker disambiguates stereo correspondence. This disambiguation is very simple, yet it yields accurate results. The current paper first quickly describes a model for underwater image formation, in the context of our recovery problem. The model accounts for flicker, scattering effects along the line of sight (LOS) and stereo formulation. The model is significantly simplified using several approximations, further described in [18]. A conclusion from this simplification is that temporal variations of flicker can establish unambiguous correspondence by local (even pointwise) calculations. The approach we present is sufficiently simple and robust to field conditions. This is demonstrated in experiments done in different locations: a swimming pool, the Red Sea and the Mediterranean.

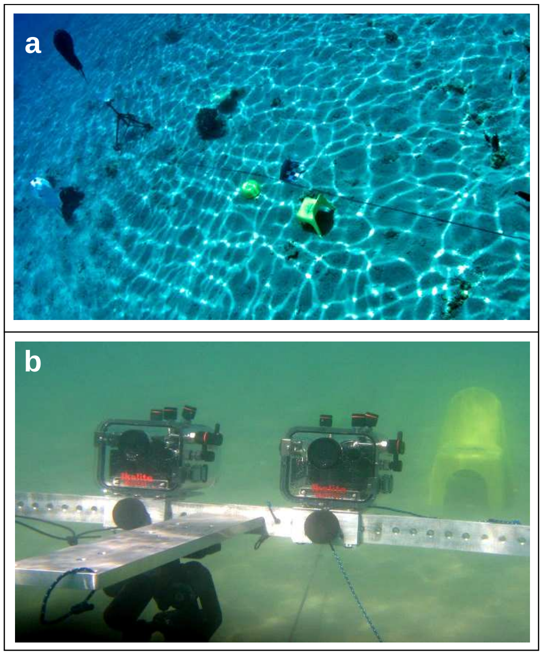

Fig. 1. [a] Sunlight flicker in the Red Sea. [b] An underwater stereoscopic video setup in the Mediterranean.

II. Theoretical Background

Denote by $\mathbf{x} = (x, y)$ an image coordinate, which corresponds to a specific LOS in the scene. Let the object radiance around a point be $I^{\text{obj}}(\mathbf{x})$. Due to attenuation in the water, the

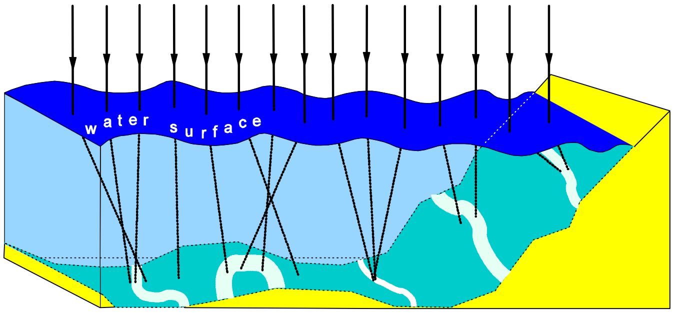

Fig. 2. A wavy water surface refracts natural illumination in a spatially varying way. Underwater, the refracted rays create a spatial pattern of lighting. The pattern varies temporally, due to the motion of the surface waves [17].

signal originating from this object [11] is

$$S(\mathbf{x}) = I^{\text{obj}}(\mathbf{x})e^{-\eta z(\mathbf{x})} \quad , \tag{1}$$

where η is the attenuation coefficient of the water. Here $z(\mathbf{x})$ is the distance between the object at $\mathbf{x}$ and the camera.

In addition to the object signal, the camera also captures *veiling light*, which is caused by ambient illumination scattered into the LOS by the water. This component is also termed *backscatter* [11], [19], and is denoted by $B(\mathbf{x})$. It is given [11] by an integral over the LOS. Overall, the radiance measured by the camera is

$$I(\mathbf{x}) = S(\mathbf{x}) + B(\mathbf{x}) \quad . \tag{2}$$

The scene illumination is not constant. In every single moment, the water surface is generally not flat but rather wavy [20], [21]. Concave and convex regions on the surface diverge and converge light rays that refract into the water.[1] This creates inhomogeneous lighting, as illustrated in Fig. 2. Consequently, the sea floor and other underwater objects are irradiated by a pattern termed *caustic networks* [24]. Due to the natural motion and evolution of the surface waves, the spatial illumination pattern changes in time, and is known as *sunlight flicker* [16]. Consequently, the irradiance in the water changes as a function of space and time. It is denoted by $I^{\text{lighting}}(\mathbf{x}, z, t)$, where t is the temporal frame index.

III. Model of Temporal Consistency

The image formation model is simple. It combines the effects described in Sec. II. We formulate the model in a manner consistent with stereo.

We use stereoscopic vision (Fig. 3). Denote the left camera by $\mathtt{L}$. We align the global coordinate system with this camera, i.e, the position of a point in the water volume or an object is uniquely defined by the left spatial coordinate vector $\mathbf{x}_{\mathtt{L}}$ and the distance z from the housing of the left camera. The right camera is denoted by $\mathtt{R}$. The object corresponding to $(\mathbf{x}_{\mathtt{L}}, z)$ in the left camera is projected to pixel $\mathbf{x}_{\mathtt{R}}$ in the right camera. The corresponding disparity vector is

$$\mathbf{d} = \mathbf{x}_{\mathtt{R}} - \mathbf{x}_{\mathtt{L}} \quad . \tag{3}$$

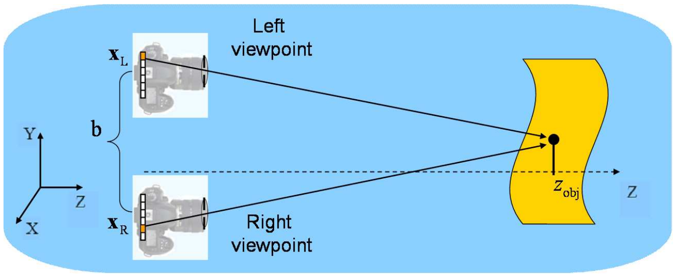

Fig. 3. A stereoscopic pair. The variables are detailed in the text.

The viewpoints of the two cameras are different, separated by a baseline of length b.

At the left camera, the signal corresponding to Eq. (1) is

$$S_{\mathtt{L}}(\mathbf{x}_{\mathtt{L}}, t) = I^{\text{lighting}}(\mathbf{x}_{\mathtt{L}}, z, t) r_{\mathtt{L}}(\mathbf{x}_{\mathtt{L}}) e^{-\eta z(\mathbf{x}_{\mathtt{L}})} \quad , \tag{4}$$

where $r_{\mathtt{L}}$ denotes the reflectance coefficient of the object towards the left camera. Eq. (4) encapsulates the temporal variations of the lighting.

Ref. [18] lists a couple of approximations that are common in stereo formulation, particularly, underwater: first, *brightness constancy* implies that $r_{\mathtt{L}} \approx r_{\mathtt{R}}$, where $r_{\mathtt{R}}$ denotes the coefficient of reflectance by the object towards the right camera. Second, the distance between the object and the right camera is sufficiently similar to the distance between the object and the left camera. Hence, the water creates a similar attenuation at both viewpoints. Under these approximations,

$$S_{\mathtt{L}}(\mathbf{x}_{\mathtt{L}}, t) \approx S_{\mathtt{R}}(\mathbf{x}_{\mathtt{R}}, t) \qquad \forall \;\; t \quad , \tag{5}$$

where $S_{\mathtt{R}}(\mathbf{x}_{\mathtt{R}}, t)$ is the signal captured by the right camera.

The spatiotemporal variations of I^{lighting} also affect the backscatter $B_{\mathtt{L}}(\mathbf{x}, t)$ and $B_{\mathtt{R}}(\mathbf{x}, t)$ in each respective camera. The relation between I^{lighting} and $B(\mathbf{x}, t)$ is more involved than Eq. (4). However, Ref. [18] shows that under common conditions, $B_{\mathtt{L}}(\mathbf{x}, t) \approx B_{\mathtt{R}}(\mathbf{x}, t)$. Compounding this result with Eqs. (2,5), the overall scene radiance, as measured by the two stereo cameras can be formulated as

$$I_{\mathtt{R}}(\mathbf{x}_{\mathtt{R}}, t) \approx I_{\mathtt{L}}(\mathbf{x}_{\mathtt{L}}, t) \qquad \forall \;\; t \quad . \tag{6}$$

Eq. (6) indicates that although the illumination varies strongly in space and time due to flicker, there is mutual consistency between the left and right viewpoints, most of the time.

IV. Correspondence From Flicker

Equation (6) claims intensity similarity at points $\mathbf{x}_{\mathtt{R}}$ and $\mathbf{x}_{\mathtt{L}}$ at time t. However, this similarity is generally not unique. A set of pixels $\Omega_{\mathtt{R}}(t) = \{\mathbf{x}_{\mathtt{R}}^{\text{incorrect}}\}$ in $I_{\mathtt{R}}$ have intensities that are very close to, or equal to $I_{\mathtt{L}}(\mathbf{x}_{\mathtt{L}})$. One reason for this is that objects at such non-corresponding pixels may have the same reflectance, irradiance and backscatter. This situation leads to the classic correspondence problem in non-flickering environments. A more general reason is that the reflectance, irradiance and backscatter in each $\mathbf{x}_{\mathtt{R}}^{\text{incorrect}}$ are all different than the ones in $\mathbf{x}_{\mathtt{R}}$, but their combination in Eq. (2) yields the same overall intensity, at time t.

[1] The wavy water surface also refracts lines of sight passing through the water surface, as described in [22], [23].

Fortunately, in flicker, such ambiguities are completely resolved with high probability, since the lighting is dynamic.[2] Due to the lighting dynamics, non-corresponding pixels in $\Omega_{\text{R}}(t)$ are generally different than those at $\Omega_{\text{R}}(t')$, at time $t' \neq t$. A coincidence of matching intensities at t has rare chances of re-occurring at t'. Considering a large number of frames N_F,

$$\bigcap_{t=1}^{N_F} \Omega_{\text{R}}(t) \longrightarrow \emptyset \ , \tag{7}$$

where in practice, even a small N_F suffices to eliminate the non-corresponding pixels.

A. Temporal Correlation

Practically, correspondence is solved in our work using mainly simple *temporal* normalized correlation. Define the vector

$$\mathbf{I}_{\text{L}}(\mathbf{x}_{\text{L}}) \equiv \begin{bmatrix} I_{\text{L}}(\mathbf{x}_{\text{L}},1) \\ I_{\text{L}}(\mathbf{x}_{\text{L}},2) \\ \vdots \\ I_{\text{L}}(\mathbf{x}_{\text{L}},N_F) \end{bmatrix} \ . \tag{8}$$

Now, in the right image, there is a set of pixels Ψ, each of which is a candidate for correspondence with $\mathbf{x}_{\text{L}}$. Without calibration of the stereo setup, Ψ is the whole field of view (all the pixels in the image). If calibration of the system had been done, then Ψ is the epipolar line [26], [27] corresponding to $\mathbf{x}_{\text{L}}$. For a candidate pixel $\mathbf{x}_{\text{R}}^{\text{cand}} \in \Psi$, define

$$\mathbf{I}_{\text{R}}(\mathbf{x}_{\text{R}}^{\text{cand}}) \equiv \begin{bmatrix} I_{\text{R}}(\mathbf{x}_{\text{R}}^{\text{cand}},1) \\ I_{\text{R}}(\mathbf{x}_{\text{R}}^{\text{cand}},2) \\ \vdots \\ I_{\text{R}}(\mathbf{x}_{\text{R}}^{\text{cand}},N_F) \end{bmatrix} \ . \tag{9}$$

Subtracting the mean of each vector, we obtain

$$\tilde{\mathbf{I}}_{\text{L}}(\mathbf{x}_{\text{L}}) = \mathbf{I}_{\text{L}}(\mathbf{x}_{\text{L}}) - \langle \mathbf{I}_{\text{L}}(\mathbf{x}_{\text{L}}) \rangle, \tag{10}$$

$$\tilde{\mathbf{I}}_{\text{R}}(\mathbf{x}_{\text{R}}^{\text{cand}}) = \mathbf{I}_{\text{R}}(\mathbf{x}_{\text{R}}^{\text{cand}}) - \langle \mathbf{I}_{\text{R}}(\mathbf{x}_{\text{R}}^{\text{cand}}) \rangle. \tag{11}$$

The empirical normalized correlation [28] between $\mathbf{x}_{\text{L}}$ and $\mathbf{x}_{\text{R}}^{\text{cand}}$ is

$$C(\mathbf{x}_{\text{R}}^{\text{cand}}) = \frac{\tilde{\mathbf{I}}_{\text{L}}(\mathbf{x}_{\text{L}})^T \tilde{\mathbf{I}}_{\text{R}}(\mathbf{x}_{\text{R}}^{\text{cand}})}{\|\tilde{\mathbf{I}}_{\text{L}}(\mathbf{x}_{\text{L}})\|_2 \|\tilde{\mathbf{I}}_{\text{R}}(\mathbf{x}_{\text{R}}^{\text{cand}})\|_2}, \tag{12}$$

where T denotes transposition. For pixel $\mathbf{x}_{\text{L}}$ in the left image, the corresponding pixel in the right image is then estimated as

$$\hat{\mathbf{x}}_{\text{R}} = \arg \max_{\mathbf{x}_{\text{R}}^{\text{cand}} \in \Psi} C(\mathbf{x}_{\text{R}}^{\text{cand}}). \tag{13}$$

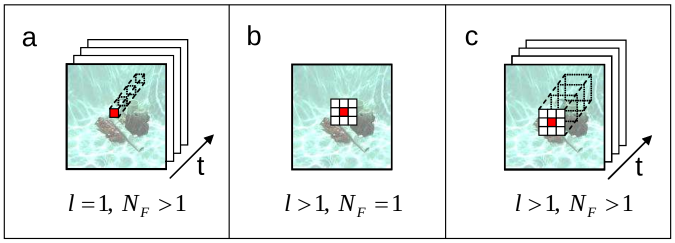

Fig. 4. Support used for [a] temporal, [b] spatial and [c] spatiotemporal correlation [18].

B. Spatiotemporal Correlation

The method described in Sec. IV-A is very simple. It does not blur range edges, since it involves no spatial operations. However, it requires correlation to be established over a large number of frames (and thus time). The number of frame can be decreased by using spatiotemporal, rather than just temporal correlation, as explained in this section. Here, the comparison is not pixel-wise, but using spatial blocks. This enables the use of smaller N_F, at the price of loss of spatial resolution and consequent range errors, particulary at range edges. Fig. 4 illustrates possibilities of correlation support.

Let $\beta(\mathbf{x}_{\text{L}})$ be a block of $l \times l$ pixels centered around $\mathbf{x}_{\text{L}}$. The pixel values in this block change during the N_F frames. Thus, the video data cube corresponding to these pixels has dimensions of $l \times l \times N_F$. Concatenate this video data cube into a vector

$$\mathbf{I}_{\text{L}}^{\text{cube}}(\mathbf{x}_{\text{L}}) \equiv [\mathbf{I}_{\text{L}}(\mathbf{x}_1)^T,\ \mathbf{I}_{\text{L}}(\mathbf{x}_2)^T, \ldots \mathbf{I}_{\text{L}}(\mathbf{x}_{l^2})]^T \ , \tag{14}$$

where $\{\mathbf{x}_m\}_{m=1}^{l^2} \in \beta(\mathbf{x}_{\text{L}})$. Analogously, an $l \times l$ block $\beta(\mathbf{x}_{\text{R}})$ is centered around $\mathbf{x}_{\text{R}}^{\text{cand}}$. Use the same concatenation as in Eq. (14) over the video in $\beta(\mathbf{x}_{\text{R}}^{\text{cand}})$ of the right camera. This yields

$$\mathbf{I}_{\text{R}}^{\text{cube}}(\mathbf{x}_{\text{R}}^{\text{cand}}) \equiv [\mathbf{I}_{\text{R}}(\mathbf{y}_1)^T,\ \mathbf{I}_{\text{R}}(\mathbf{y}_2)^T, \ldots \mathbf{I}_{\text{R}}(\mathbf{y}_{l^2})]^T \ , \tag{15}$$

where $\{\mathbf{y}_m\}_{m=1}^{l^2} \in \beta(\mathbf{x}_{\text{R}}^{\text{cand}})$.

Now, Eqs. (10,11) are redefined as

$$\tilde{\mathbf{I}}_{\text{L}}(\mathbf{x}_{\text{L}}) = \mathbf{I}_{\text{L}}^{\text{cube}}(\mathbf{x}_{\text{L}}) - \langle \mathbf{I}_{\text{L}}^{\text{cube}}(\mathbf{x}_{\text{L}}) \rangle, \tag{16}$$

$$\tilde{\mathbf{I}}_{\text{R}}(\mathbf{x}_{\text{R}}^{\text{cand}}) = \mathbf{I}_{\text{R}}^{\text{cube}}(\mathbf{x}_{\text{R}}^{\text{cand}}) - \langle \mathbf{I}_{\text{R}}^{\text{cube}}(\mathbf{x}_{\text{R}}^{\text{cand}}) \rangle. \tag{17}$$

Eqs. (16,17) are then used in Eqs. (12,13).

Spatial Correlation: A degenerate case is to use a single stereo frame-pair, i.e, $N_F = 1$, while $l > 1$. Here, only spatial correlation is performed. This is a common stereo practice [27], [29]. Matching that is based solely on spatial correlation requires significant spatial texture. Thus, the spatial variations in the caustic lighting field provides some texture over areas having textureless albedo. This had been used [12] underwater to enhance the correspondence, independently per each individual stereo frame-pair.

V. Reliable and Unreliable Results

There are situations in which the technique is unreliable. Fortunately, such problems can often be predicted. Some pixels simply correspond to object points that reside in the shadow

[2] In Ref. [25], man-made light patterns illuminate the scene using a projector in order to establish correspondence between stereo video cameras. In scattering media, artificial illumination is problematic, since it cannot irradiate distant objects [11]. Artificial structured illumination is often designed to be narrow, to reduce excessive backscatter [19]. In our case, lighting variations are natural and are anywhere along the LOS.

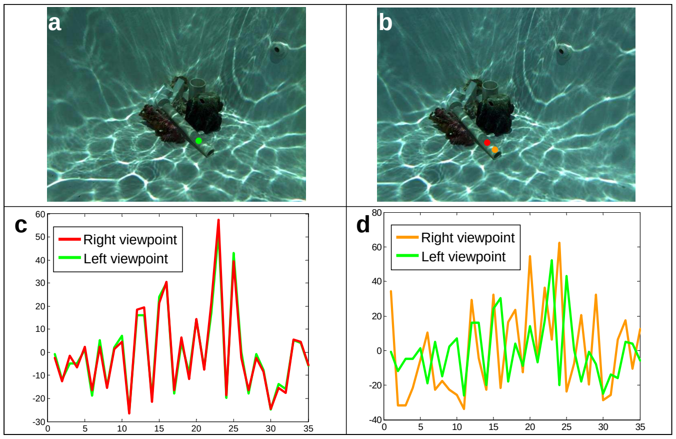

Fig. 5. Left [a] and Right [b] frames at one instance in the sequence. [c] Temporal plots of $\tilde{\mathbf{I}}_{\rm L}(\mathbf{x}_{\rm L})$ and $\tilde{\mathbf{I}}_{\rm R}(\hat{\mathbf{x}}_{\rm R})$ extracted from corresponding pixels. These pixels are marked by green and red in the respective frames [a,b]. [d] Temporal plots of $\tilde{\mathbf{I}}_{\rm L}(\mathbf{x}_{\rm L})$ and $\tilde{\mathbf{I}}_{\rm R}(\hat{\mathbf{x}}_{\rm R})$ extracted from non-corresponding pixels. These pixels are marked by green and orange in the respective frames [a,b].

of downwelling lighting, due to objects above them. Points in the shadow are unaffected by flicker. Similarly, for objects which are very far away, the signal is attenuated (Eq. 4), thus it is difficult to sense the temporal variations due to flicker there in short periods. In a pixel corresponding to such problematic cases, the temporal standard deviation of the pixel value $\|\tilde{\mathbf{I}}_{\rm L}(\mathbf{x}_{\rm L})\|_2$ is very low. Hence, the set of low-signal pixels can be assessed by thresholding the field $\|\tilde{\mathbf{I}}_{\rm L}(\mathbf{x}_{\rm L})\|_2$ by a parameter $\tau_{\rm STD}$.

Correspondence of an object point is impossible if the point is occluded at a viewpoint. The set of pixels in $\mathbf{I}_{\rm L}$ that are occluded in $\mathbf{I}_{\rm R}$ have a low value of C even in the "optimal" match $\hat{\mathbf{x}}_{\rm R}$. Hence, if $C(\hat{\mathbf{x}}_{\rm R})$ is below a threshold τ_C, it indicates an unreliable correspondence. Thus, Ref. [18] defined a set ρ of reliable pixels by

$$\rho = \{\mathbf{x}_{\rm L} :\ [C_{\mathbf{x}_{\rm L}} > \tau_C] \text{ AND } [\|\tilde{\mathbf{I}}_{\rm L}(\mathbf{x}_{\rm L})\|_2 > \tau_{\rm STD}]\}\ . \qquad (18)$$

VI. Experiments

We conducted a set of in-situ field experiments. Different scenes and cameras were used, in the ocean and in a pool. In this section, the results of these experiments are shown and discussed.

A. Swimming-Pool Experiment

Consider our first example, which is an experiment conducted in a swimming pool. The scene includes several objects at $z \in [1\text{m}, 2\text{m}]$, near the corner of the pool. The depth at the bottom of the pool was ~ 1m. The stereo setup was a *Videre Design* [30] head shooting at 7fps, with $b = 25$cm. A sample frame-pair appears in Fig. 5a,b. Temporal correlation was performed using $N_F = 35$. Here, as in all the following experiments, the setup was not calibrated, hence the search domain Ψ includes the entire field of view.[3] Examples of temporal matches in corresponding and non-corresponding points are shown in Figs. 5c and 5d, respectively.

The results of temporal correlation are shown in Fig. 6. As common in studies dealing with stereo correspondence [31], the result is displayed as a *disparity map*, rather than a range map.[4] The disparity map is derived based on Eq. (3):

$$\hat{d}(\mathbf{x}_{\rm L}) = \|\hat{\mathbf{x}}_{\rm R} - \mathbf{x}_{\rm L}\|\ . \qquad (19)$$

The disparity map of the pool experiment is shown in Fig. 6b.

It may be noticed that there are a few small regions with clearly outlying results. These regions were in constant shadow, hence without any flicker. This is discussed in Sec. V.

In this experiment, water visibility was good. This allowed us to extract quantitative performance measures based on manual matching. In the field of view, 100 points were randomly selected in $I_{\rm L}$. These points were manually matched in $I_{\rm R}$. This match served as ground truth in the tests. First, Fig. 7 plots the required N_F as a function of the required reliability of matching, where epipolar constraints were not put to use.

Then, using the same video data, we re-ran the recovery using spatiotemporal correlation, as described in Sec. IV-B, using various values of l. Qualitatively, the resulting disparity maps resemble those of Fig. 6. The quantitative plots in Fig. 7, however, show that with large spatial support, a moderate success rate of $\approx 80-85\%$ can be achieved using much fewer frames than if using only temporal correlation. However, widening the spatial support stagnates the success rate below $\approx 90\%$ even when the number of frames grows. Possibly, this is caused by true violations of spatial smoothness in the range map. In contrast, in sole pointwise analysis, the success rate increases monotonically with time and eventually surpasses the results achieved using spatial matching windows.

As mentioned in Sec. IV-B, correspondence may also be sought using only spatial correlation, in a single stereo pair ($N_F = 1$) of a flicker scene. It was shown in [18] that using flicker spatial variation, the success rate was $\approx 60\%$. However, after filtering out flicker, the spatial correlation matched only 17% of the points. This indicates that flicker spatial information is also valuable for stereo correspondence.

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. Photographs of the stereo setup are shown in Figs. 1 and 8.

Here, we used Canon HV-30 high-definition PAL video cameras within Ikelite underwater housings. To synchronize the video sequences, blinking flashlight was shined into the

[3] Since epipolar geometry was not exploited to limit the match search, a few erroneous matches appeared, which would have been bypassed with epipolar search constraints. These singular errors were effectively eliminated from the disparity map using a 3×3 median filter.

[4] A range map can be derived from the correspondences, once the setup is calibrated. Underwater, such a calibration may not match the calibration model in air, since the water interface introduces a non-single viewpoint geometry [32] to each camera.

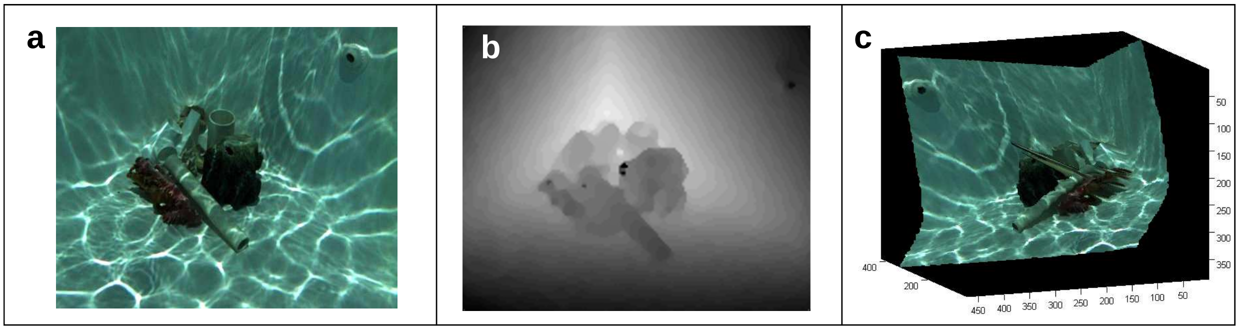

Fig. 6. [a] A frame from the left viewpoint in the pool experiment. The estimated disparity map $\|\hat{d}\|$ is shown in [b]. Its reciprocal, which is similar to the range map, is used for texture mapping a different viewpoint in [c].

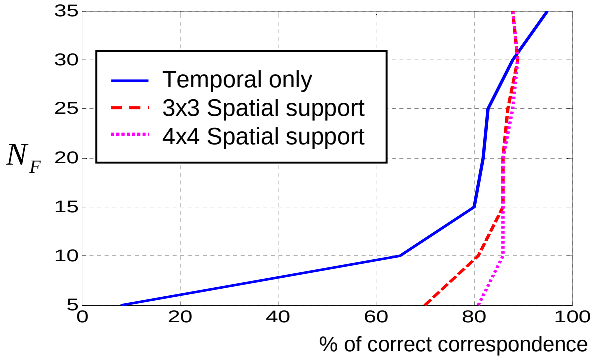

Fig. 7. The number of frames required to achieve a certain rate of successful match in the experiment corresponding to Fig. 6 [18].

running cameras before and after each experiment. These blinks were later detected in postprocessing 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.

One experiment conducted in the Red Sea is shown in Fig. 9. Here, visibility was better than in the Mediterranean experiments. The baseline was $b = 30$cm and $N_F = 50$. The distance of the bowl, the board and the chair was 2m, 2.5m and 3m respectively.

In another experiment, done in a different day, 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. 10. The distance of the large cube from the cameras was ~ 5m.

Another oceanic experiment is depicted in Fig. 11. Here visibility was very poor, leading to shorter objects distances. The distance of the chair was 2m. Consequently the baseline was short ($b = 30$cm) and $N_F = 75$.

As explained in Sec. V, there is an automatic determination of pixels having low reliability of the match. Such pixels are marked in black in Figs. 10 and 11. They appear mainly in shadowed areas or occluded regions in the right viewpoint.

VII. Conclusions

Natural underwater flicker is helpful. It leads to a dense correspondence map, using video. We believe that this approach can be a basis for a wide range of shallow water engineering applications, such as mapping, archaeology, navigation and inspection of boats, structures and pipes. The method establishes correspondence rather reliably even without epipolar constraints. Hence, we hypothesize that a sequence of such correspondence mappings can possibly establish the epipolar geometry of the system.

Acknowledgements

We thank Izchak Yogev, Tali Treibitz, Amin Sarafraz and Ali Taatian for their help in conducting the underwater experiments. 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 (BSF) Grant 2006384. This work was conducted in the Ollendorff Minerva Center. Minerva is funded through the BMBF.

References

[1] T. Boult, “DOVE: Dolphin omni-directional video equipment,” in *Proc. IASTED Int. Conf. Robotics and Autom.*, 2000, pp. 214–220.

[2] D. M. Kocak, F. R. Dalgleish, F. M. Caimi, and Y. Y. Schechner, “A focus on recent developments and trends in underwater imaging,” *MTS J.*, vol. 42, no. 1, pp. 52–67, 2008.

[3] A. Sarafraz, S. Negahdaripour, and Y. Y. Schechner, “Enhancing images in scattering media utilizing stereovision and polarization,” *In Proc. IEEE WACV*, 2009.

[4] M. Bryant, D. Wettergreen, S. Abdallah, and A. Zelinsky, “Robust camera calibration for an autonomous underwater vehicle,” in *Proc. Australian Conf. on Robotics and Autom.*, 2000, pp. 111–116.

[5] J. Sattar and G. Dudek, “Where is your dive buddy: tracking humans underwater using spatio-temporal features,” in *In Proc. IEEE/RSJ IROS*, 2007.

[6] G. L. Foresti, “Visual inspection of sea bottom structures by an autonomous underwater vehicle,” *IEEE Trans. Syst. Man and Cyber*, vol. 31, pp. 691–705, 2001.

[7] Y. Kahanov and J. Royal, “Analysis of hull remains of the Dor D vessel, Tantura lagoon, Israel,” *Int. J. Nautical Archeology*, vol. 30, pp. 257–265, 2001.

[8] T. W. Cronin and J. Marshall, “Parallel processing and image analysis in the eyes of mantis shrimps,” *Biol. Bull.*, vol. 200, pp. 177–183, 2001.

[9] 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.*, vol. 138, pp. 155–179, 1988.

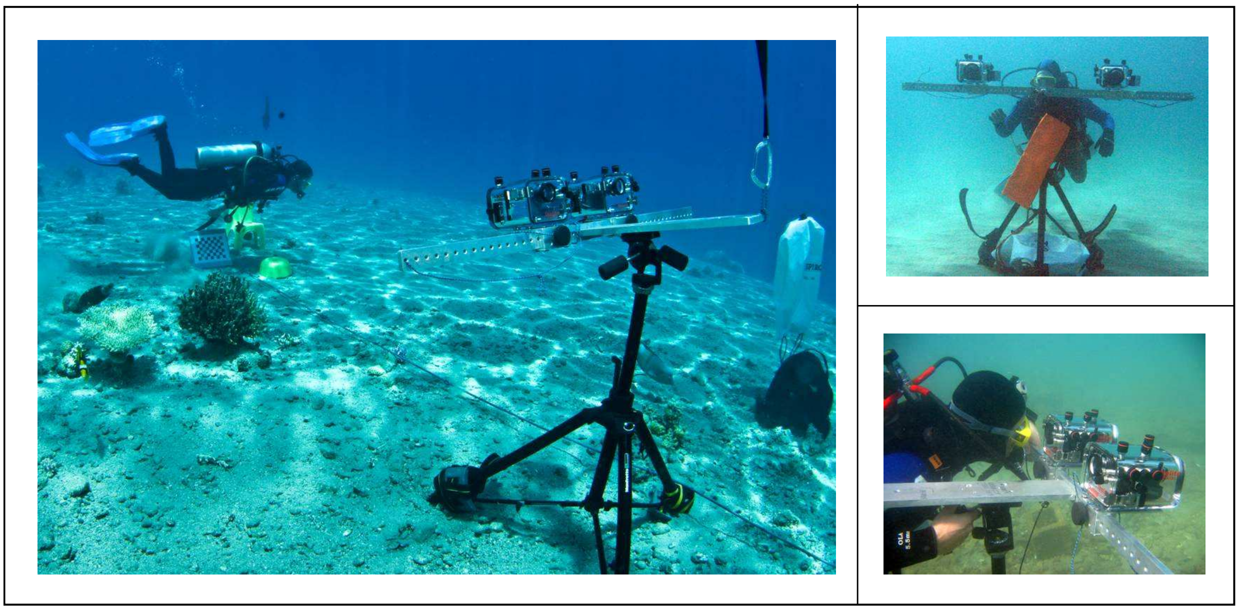

Fig. 8. The oceanic experiment setup. Here, $b \in [30, 70]$ cm.

[10] J. Marshall, T. W. Cronin, and S. Kleinlogel, "Stomatopod eye structure and function: A review," *Arthropod Structure and Development*, vol. 36, pp. 420–448, 2007.

[11] Y. Y. Schechner and N. Karpel, "Recovery of underwater visibility and structure by polarization analysis," *IEEE J. Oceanic Eng.*, vol. 30, pp. 570–587, 2005.

[12] S. Negahdaripour and P. Firoozfam, "An ROV stereovision system for ship-hull inspection," *IEEE J. Oceanic Eng.*, vol. 31, pp. 551–564, 2006.

[13] J. M. Lavest, F. Guichard, and C. Rousseau, "Multiview reconstruction combining underwater and air sensors," in *Proc. IEEE ICIP.*, vol. 3, 2002, pp. 813–816.

[14] N. G. Jerlov, *Marine Optics*. Elsevier, Amsterdam, 1976, ch. 6.

[15] R. E. Walker, *Marine Light Field Statistics*. John Wiley, New York, 1994, ch. 10.

[16] 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*, 2008.

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

[18] Y. Swirski, Y. Y. Schechner, B. Herzberg, and S. Negahdaripour, "Stereo from Flickering Caustics," *In Proc. IEEE ICCV*, 2009.

[19] M. Gupta, S. Narasimhan, and Y. Y. Schechner, "On controlling light transport in poor visibility environments," in *Proc. IEEE CVPR*, 2008.

[20] A. Fournier and W. T. Reeves, "A simple model of ocean waves," in *Proc. SIGGRAPH*, 1986, pp. 75–84.

[21] M. Gamito and F. Musgrave, "An accurate model of wave refraction over shallow water," *Computers and Graphics*, vol. 26, pp. 291–307, 2002.

[22] A. A. Efros, V. Isler, J. Shi, and M. Visontai, "Seeing through water," in *Proc. NIPS 17*, 2004, pp. 393–400.

[23] Y. Tian and S. G. Narasimhan, "Seeing through water: Image restoration using model-based tracking," *In Proc. IEEE ICCV*, 2009.

[24] D. K. Lynch and W. Livingston, *Color and Light in Nature*, 2nd ed. Cambridge U.Press, 2001, ch. 2.4,2.5,3.7,3.16.

[25] J. Davis, D. Nehab, R. Ramamoorthi, and S. Rusinkiewicz, "Spacetime stereo: A unifying framework for depth from triangulation," *IEEE Trans. PAMI*, vol. 27, pp. 296–302, 2005.

[26] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003, ch. 9-12.

[27] E. Trucco and A. Verri, *Introductory Techniques For 3D Computer Vision*. Prentice Hall, New Jersey, 1998, ch. 6.

[28] R. Eustice, O. Pizarro, H. Singh, and J. Howland, "UWIT: underwater image toolbox for optical image processing and mosaicking in Matlab," in *Proc. Int. Sympos. on Underwater Tech.*, 2002, pp. 141– 145.

[29] R. Bolles, H. Baker, and M. Hannah, "The JISCT stereo evaluation," in *Proc. DARPA Image Understanding Workshop*, 1993, pp. 263–274.

[30] G. C. Boynton and K. J. Voss, "An underwater digital stereo video camera for fish population assessment," University of Miami, Tech. Rep., 2006.

[31] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *IJCV*, vol. 47, pp. 7–42, 2002.

[32] T. Treibitz, Y. Y. Schechner, and H. Singh, "Flat refractive geometry," *In Proc. IEEE CVPR*, 2008.

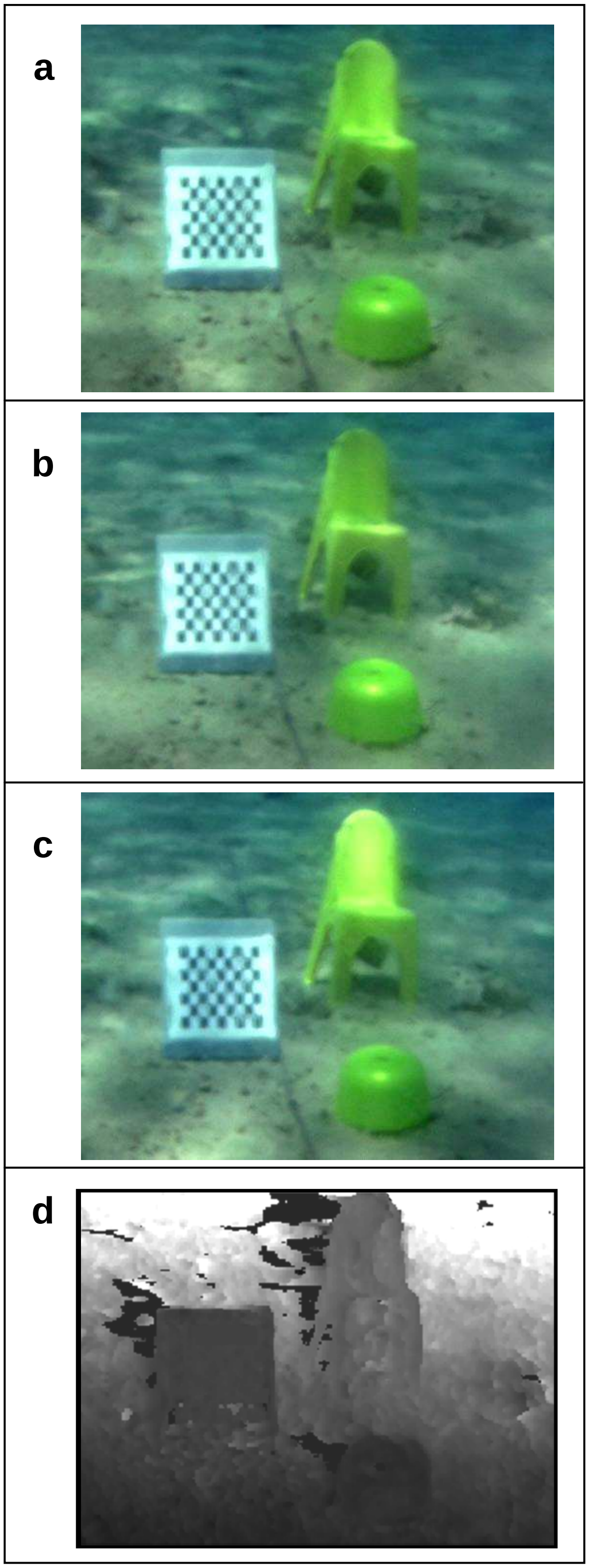

Fig. 9. [a-c] Raw left frames from the Red Sea experiment. [d] The estimated disparity map. Black areas represent low correspondence reliability.

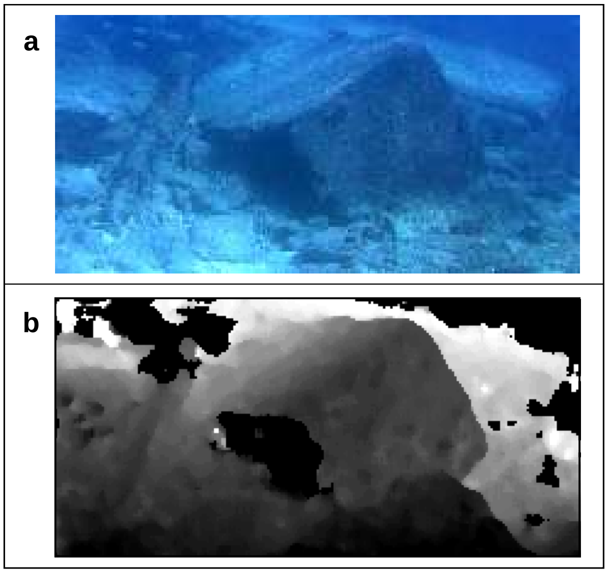

Fig. 10. [a] A raw left frame from an experiment in a marine archaeological site (Caesarea). [b] The estimated disparity map. Black areas represent low correspondence reliability.

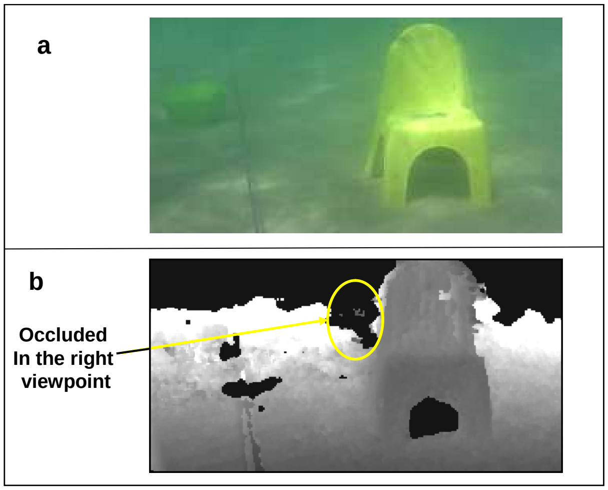

Fig. 11. [a] A raw left frame from a second oceanic experiment. [b] The estimated disparity map. Black areas represent low correspondence reliability [18].