

Variational Stereo in Dynamic Illumination

Yohay Swirski, Yoav Y. Schechner*
 Dept. of Electrical Engineering
 Technion - Israel Inst. Technology
 Haifa 32000, Israel

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

Tal Nir
 Dept. of Computer Science
 Technion - Israel Inst. Technology
 Haifa 32000, Israel

taln@cs.technion.ac.il

Abstract

Temporal irradiance variations are useful for finding dense stereo correspondences. These variations can be created artificially using structured light. They also occur naturally underwater. We introduce a variational optimization formulation for finding a dense stereo correspondence field. It is based on multi-frame optical flow, adapted to stereo. The formulation uses a sequence of stereo frames, and yields dense and robust results. The inherent aperture problem of optical flow is resolved using a temporal sequence of stereo frame-pairs. The results are achieved even without considering epi-polar geometry. The method has the ability to handle dynamic stereo underwater, in harsh conditions of flickering illumination. The method is demonstrated experimentally both outdoors and indoors.

1. Introduction

Stereo correspondence is well-studied in computer vision [25]. An important set of methods for establishing correspondence uses optimization methods: graph cuts [5], belief propagation [30] and methods based on an optical flow [6, 14, 29] formulation. Dense correspondence achievable using optical flow formulation for stereo systems is useful to refine [1] or calibrate [19, 37] the epipolar geometry. Furthermore, an optical flow formulation partly compensates for lack of local constraints, using flow smoothness terms. However, the accuracy and reliability of methods seeking dense correspondence fields eventually depend on the scene texture.

Irrespective of optimization formulation for stereo, Refs. [9, 40] show that spatiotemporal information can be useful for finding dense correspondence, using structured light. Spatiotemporal variations also occur naturally underwater (Fig. 1.a). There, an effect called *sunlight flicker*¹ [12, 28] exists. Submerged objects are illuminated

*This work relates to 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.

¹Interestingly, in many marine animals, vision is adapted to spatiotem-

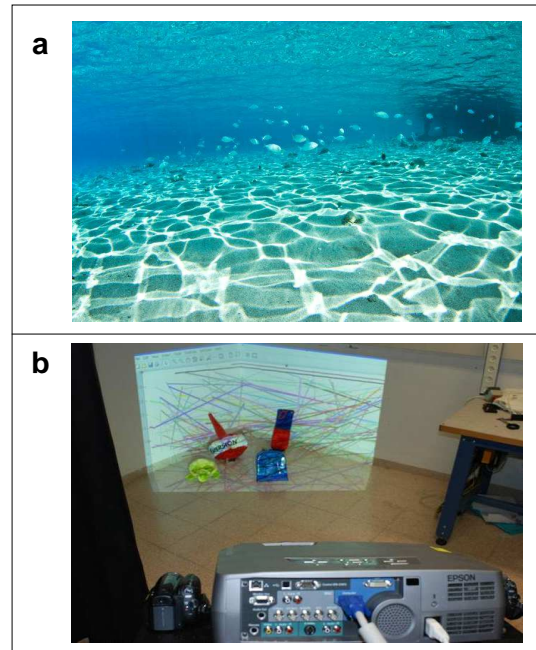


Figure 1. [a] Sunlight flicker irradiating a scene in the sea. Courtesy of Itzhak Yogev. [b] A projected spatiotemporal varying pattern on an indoor scene.

by a natural random pattern (caustics) [10, 38]. The refraction of sunlight through a wavy water surface creates inhomogeneous lighting. This has implications for underwater computer vision. The domain of underwater computer vision [4, 17, 27, 36] focuses on oceanic engineering, which includes automated vehicle control [7, 15], inspection of pipelines [11], communication cables, ports and ship hulls [23]. Computer vision is also used inside swimming pools [18]. Refs. [31, 32, 33] show that sunlight flicker can be useful for finding dense correspondence in underwater stereo, using spatiotemporal correlation. However, a relatively large number of frames is needed in order to acquire enough local constraints to disambiguate correspondence.

poral frequencies of flicker [20].

This paper combines *optical flow for stereo* and *spatiotemporal stereo* into a unified formulation. The formulation uses variational expressions for dense correspondence in stereoscopic video, where spatiotemporal irradiance variations exist. The formulation uses a sequence of frames.

A scene is acquired from two viewpoints, in video. If the scene irradiance changes temporally, each stereo pair of frames adds information to a data term of an optimization cost function. This yields an accurate and dense correspondence map, with a relatively small number of frames and fast computations. This, in turn, enables *dynamic stereo*, where camera or scene motions exist. We demonstrate this feature in an underwater experiment.

2. Background: Variational Optical Flow

Let a camera take images $I(x, y, k)$ at different time samples k . The optical flow components u, v are the local warping displacements between consecutive frames. The general optical flow functional [14] is

$$\int_{\Omega} \left\{ \Psi_D [I(x+u, y+v, k+1) - I(x, y, k)]^2 + \alpha \Psi_S (u_x^2 + u_y^2 + v_x^2 + v_y^2) \right\} dx dy, \quad (1)$$

where Ω represents the entire image domain. The first term in Eq. (1) is the data term. It expresses expectation for brightness constancy. The second term is a smoothness term and α is the smoothness weight. If

$$\Psi_{D,S} (s^2) = s^2, \quad (2)$$

then $\Psi_{D,S}$ is the L_2 functional [14]. If, however,

$$\Psi_{D,S} (s^2) = \sqrt{s^2 + \varepsilon_{D,S}^2}, \quad (3)$$

then $\Psi_{D,S}$ is an approximation to the L_1 functional [6]. Here $\varepsilon_{D,S}$ are small constants which make $\Psi_{D,S}$ differentiable. The L_1 norm in the data term is robust to outliers in the images. The L_1 norm in the smoothness term avoids the over-smoothing at flow discontinuities, which may occur in L_2 smoothness term. Refs. [2, 21, 39] suggested extensions of this formulation to multi-frame scenarios. This allows using spatio-temporal smoothness instead of spatial smoothness. When an optical flow formulation is used to establish stereo correspondence [19, 29], the L_1 norm enables sharp discontinuities and occlusions in the disparity map.

3. Stereo by Spatio-Temporal Optical Flow

3.1. Motivation for Using a Sequence of Pairs

The use of several pairs of frames for optical flow estimation has several advantages, particularly overcoming the

aperture problem. Let a static scene be captured by N_F pairs of frames, from two viewpoints. Each pair is inter-related by an optical flow field. The linearized brightness constancy equation of the k^{th} frame is

$$I_x^k u + I_y^k v + I_{LR}^k = 0, \quad (4)$$

where I_x^k and I_y^k are the partial spatial derivatives of the k^{th} frame. In the classical optical flow problem, I_{LR}^k is the change in brightness over time. In stereo, I_{LR}^k is the change in brightness between images taken simultaneously from two viewpoints. Denote by ∇I_k^T the image gradient in the k^{th} frame. Re-writing Eq. (4):

$$\nabla I_k^T \begin{pmatrix} u \\ v \end{pmatrix} = -I_{LR}^k, \quad (5)$$

where T denotes transposition. Eq. (5) implies that locally, only the flow component in the direction of the image gradient can be determined. This is the well-known aperture problem.

Recall that different pairs of frames of the static scene exist. Eq. (5) applies to each pair of frames. Therefore, more constraints are available:

$$\begin{pmatrix} \nabla I_1^T \\ \nabla I_2^T \\ \vdots \\ \nabla I_{N_F}^T \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = - \begin{pmatrix} I_{LR}^1 \\ I_{LR}^2 \\ \vdots \\ I_{LR}^{N_F} \end{pmatrix}. \quad (6)$$

Considering Eq. (6), if there are at least two stereo frame-pairs, where the gradient directions are non-parallel, then the aperture problem is resolved. If the provided pairs involve significant temporal light variations, the probability that the gradient directions coincide is small, especially when three or more pairs are provided. This bypasses the aperture problem, and increases the accuracy of the computed flow, instead of relying only on regularization to fill-in constraints. This is a major cause for the significant increase in accuracy demonstrated in Sec. 4.

3.2. The Cost Function

Let I_L^k and I_R^k denote the k^{th} left and right frames. A matching functional should penalize for the differences in all the pairs of frames. Generalizing Eq. (1) to multiple frames,

$$\int_{\Omega} \left\{ \sum_{k=1}^{N_F} \Psi_D [I_L^k(x+u, y+v) - I_R^k(x, y)]^2 + \alpha \Psi_S (u_x^2 + u_y^2 + v_x^2 + v_y^2) \right\} dx dy. \quad (7)$$

We use the L_1 norm, as described in Sec. 2. Therefore, Ψ is determined according to Eq. (3) in our scheme.

3.3. Linearization

We linearize the cost function to simplify the numerical scheme. The linearization is limited to small changes in disparity, per scale. In order to enable large disparity vectors, a coarse-to-fine approach is applied, as we describe in Sec. 3.6. Linearizing $I_L^k(x+u, y+v)$ around (u_0, v_0) ,

$$I_L^k(x+u, y+v) \cong I_L^k(x+u_0, y+v_0) + I_{L,x}^k(x+u_0, y+v_0)\delta u + I_{L,y}^k(x+u_0, y+v_0)\delta v, \quad (8)$$

where $\delta u = u - u_0$ and $\delta v = v - v_0$. Here, $I_{L,x}$ and $I_{L,y}$ are the partial spatial derivatives of I_L . Substituting Eq. (8) into Eq. (7) and re-writing it, we get

$$\int_{\Omega} \left[\sum_{k=1}^{N_F} \Psi_D \left(\tilde{I}_x^k u + \tilde{I}_y^k v + \tilde{I}_{LR}^k \right)^2 + \alpha \Psi_S \left(u_x^2 + u_y^2 + v_x^2 + v_y^2 \right) \right] dx dy, \quad (9)$$

where:

$$\tilde{I}_x^k := I_{L,x}^k(x+u_0, y+v_0), \quad \tilde{I}_y^k := I_{L,y}^k(x+u_0, y+v_0) \quad (10)$$

and

$$\tilde{I}_{LR}^k := I_L^k(x+u_0, y+v_0) - I_R^k(x, y) - \tilde{I}_{L,x}^k u_0 - \tilde{I}_{L,y}^k v_0. \quad (11)$$

3.4. Numerical Scheme

The numerical scheme for solving this functional can be formulated as a re-weighted least means squares [24] functional²:

$$\int_{\Omega} \left[\sum_{k=1}^{N_F} w_D^k \left(\tilde{I}_x^k u + \tilde{I}_y^k v + \tilde{I}_{LR}^k \right)^2 + \alpha w_S \left(u_x^2 + u_y^2 + v_x^2 + v_y^2 \right) \right] dx dy, \quad (12)$$

where w_D^k is updated every n_{update} iterations at the same rate of u_0, v_0 updates. The smoothness weight w_S is updated at every iteration.

The data and smoothness weights are set to be

$$w_D^k = \frac{1}{\sqrt{[I_L^k(x+u_0, y+v_0) - I_R^k(x, y)]^2 + \varepsilon_D^2}}, \quad (13)$$

$$w_S = \frac{1}{\sqrt{u_x^2 + u_y^2 + v_x^2 + v_y^2 + \varepsilon_S^2}}. \quad (14)$$

²We use a re-weighted least mean square numerical scheme due to its simplicity and relatively fast computation. The main disadvantage of this scheme is inability to use very low values for ε . Using such low values of ε requires more complex numerical schemes, as described in [6]. We found that using non-negligible ε does not over-smooth sharp edges in the disparity map, as can be noticed in Sec. 4.

Similarly to Ref. [14], the two Euler-Lagrange equations derived from Eq. (12) are

$$\sum_{k=1}^{N_F} w_D^k \left(\tilde{I}_x^k u + \tilde{I}_y^k v + \tilde{I}_{LR}^k \right) \tilde{I}_x^k - \alpha w_S (\nabla^2 u) = 0, \quad (15)$$

$$\sum_{k=1}^{N_F} w_D^k \left(\tilde{I}_x^k u + \tilde{I}_y^k v + \tilde{I}_{LR}^k \right) \tilde{I}_y^k - \alpha w_S (\nabla^2 v) = 0. \quad (16)$$

Define:

$$\begin{aligned} A_1 &= \sum_{k=1}^{N_F} w_D^k \left(\tilde{I}_x^k \right)^2, & A_2 &= \sum_{k=1}^{N_F} w_D^k \tilde{I}_x^k \tilde{I}_y^k \\ A_3 &= \sum_{k=1}^{N_F} w_D^k \left(\tilde{I}_y^k \right)^2, & A_4 &= \sum_{k=1}^{N_F} w_D^k \tilde{I}_x^k \tilde{I}_{LR}^k \\ A_5 &= \sum_{k=1}^{N_F} w_D^k \tilde{I}_y^k \tilde{I}_{LR}^k \end{aligned} \quad (17)$$

Then, Eqs. (15,16) can be expressed as

$$A_1 u + A_2 v + A_4 - \alpha w_S (\nabla^2 u) = 0, \quad (18)$$

$$A_2 u + A_3 v + A_5 - \alpha w_S (\nabla^2 v) = 0. \quad (19)$$

Inserting a discrete approximation for the Laplacian [14] in Eqs. (18,19) yields

$$A_1 u + A_2 v + A_4 - \alpha w_S (\bar{u} - u) = 0, \quad (20)$$

$$A_2 u + A_3 v + A_5 - \alpha w_S (\bar{v} - v) = 0. \quad (21)$$

Here, \bar{u} denotes a weighted average of neighboring pixels (the same approximation is used for \bar{v}):

$$\bar{u}(x, y) = \frac{\sum_{i=-1}^1 \sum_{j=-1}^1 w_{i,j}(x, y) u(x+i, y+j)}{\sum_{i=-1}^1 \sum_{j=-1}^1 w_{i,j}(x, y)}. \quad (22)$$

Here, $w_{i,j}(x, y)$ is defined as

$$w_{i,j}(x, y) = \begin{cases} 2, & |i| + |j| = 1; \\ 1, & |i| + |j| = 2; \\ 0, & i = j = 0. \end{cases} \quad (23)$$

We shall refer to the weighting average of Eqs. (22,23) as *non-directional*. Re-arranging Eqs. (20,21), we get:

$$u = \bar{u} - \frac{A_1 \bar{u} + A_2 \bar{v} + A_4}{D}, \quad (24)$$

$$v = \bar{v} - \frac{A_2 \bar{u} + A_3 \bar{v} + A_5}{D}, \quad (25)$$

where, $D = \alpha w_S + A_1 + A_3$.

Eqs. (24,25) are linear equations of u and v . However, they contain linear convolutions, defined in Eq. (22). Therefore, we use an iterative Gauss-Seidel solver [16] to estimate u and v .

3.5. Directional Smoothness Term

The L_1 smoothness term in Eq. (3) is non-directional. Flow field discontinuities reduce the smoothness weight w_S according to Eq. (14), thus enforcing less smoothness in all directions. We wish to introduce directional smoothness, in order to selectively reduce the influence of the smoothness term and better preserve edges. This is done by replacing the constant weights $w_{i,j}(x, y)$ by

$$w_{i,j}(x, y) = \frac{1}{\sqrt{\Delta u_{i,j}(x, y)^2 + \Delta v_{i,j}(x, y)^2 + \varepsilon_S^2}}$$

$$w_{0,0}(x, y) = 0, \quad (26)$$

where

$$\Delta u_{i,j}(x, y) = u(x + i, y + j) - u(x, y). \quad (27)$$

The same weights $w_{i,j}(x, y)$ are used for \bar{u} and \bar{v} . Suppose a discontinuity in the flow field occurs between two adjacent pixels. Then, the corresponding smoothness weights of both u and v are reduced, but the smoothness weight reduction is focused on the discontinuity direction. So, when there is a discontinuity, it affects both components of the flow field, without decreasing smoothness weights in a direction where there is continuity. Eq. (26) determines the weights selectively, thus changing the strength of the smoothness bonds between neighboring pixels.

The directional smoothness term is applied by calculating \bar{u} and \bar{v} in Eq. (22) using directional weights (26) and setting $w_S = 1$. The advantages of directional weighting in the smoothness term are demonstrated in Sec. 4.

3.6. Coarse to Fine Solution

As in [6], a pyramidal coarse-to-fine approach is used. In the coarse levels, spatio-temporal variations make only a small contribution, since they are averaged by the low-pass spatial filters applied to create the coarse levels. The coarse levels enable large disparity vectors in a variational framework. At the finer scales, the contribution of temporal variations becomes significant. Therefore, the main difference between spatial-only to spatio-temporal solutions is in the fine details. This fact is supported by results shown in Sec. 4.

3.7. Normalization

The brightness constancy equation may be violated due to brightness changes between the views. This may be caused by different sensitivity, aperture and exposure time settings.³ Local brightness differences between viewpoints

³Refs. [13] and [22] suggested considering physical models of brightness changes in the optical flow formulation. Ref. [6] required gradient constancy in the data term to handle this problem.

may be created by non-Lambertian reflections and spatio-temporal varying backscatter underwater [31].

To counter these problems, we apply local brightness normalization as a pre-process, before optical flow estimation. The normalized intensity is given by:

$$I_{\text{Normalized}} = \frac{I - \text{Mean}}{\sqrt{\text{STD}^2 + \beta^2}}, \quad (28)$$

where Mean and STD respectively denote the average and standard deviation in a neighborhood. Here, β is used to moderate amplification of image noise in areas where the brightness is nearly uniform.

3.8. Summary

This section summarizes our implementation for obtaining correspondences in stereoscopic video with temporal light variations, using spatio-temporal optical flow formulation.

1. Normalize the images as explained in Sec. 3.7.
2. Construct a Gaussian pyramid of the images.
3. Starting from the lowest resolution, initialize the optical flow with zero flow.
4. Calculate the weights in Eq. (13), \bar{u} , \bar{v} and the summations in Eq. (17). We use a bilinear interpolation for sub-pixel sampling.
5. Use Eqs. (24,25) to update the flow field for several iterations.
6. Repeat steps 4 and 5 several times until the optical flow estimation converges.
7. Expand the optical flow estimation to the finer scale level.
8. Repeat steps 4-7 until the optical flow is estimated at the finest scale.

4. Experiments

4.1. Parameter Settings

This section describes the empirical parameter settings in our experiments. The same set of parameters is used for all the sequences and tests. The smoothness weight we found preferable is

$$\alpha = 20N_F. \quad (29)$$

It increases linearly with N_F to maintain the same relative weight between data and smoothness terms.⁴ For the L_1

⁴When N_F is large, α can be reduced, since there may be sufficient data to determine accurate correspondence with little regularization.

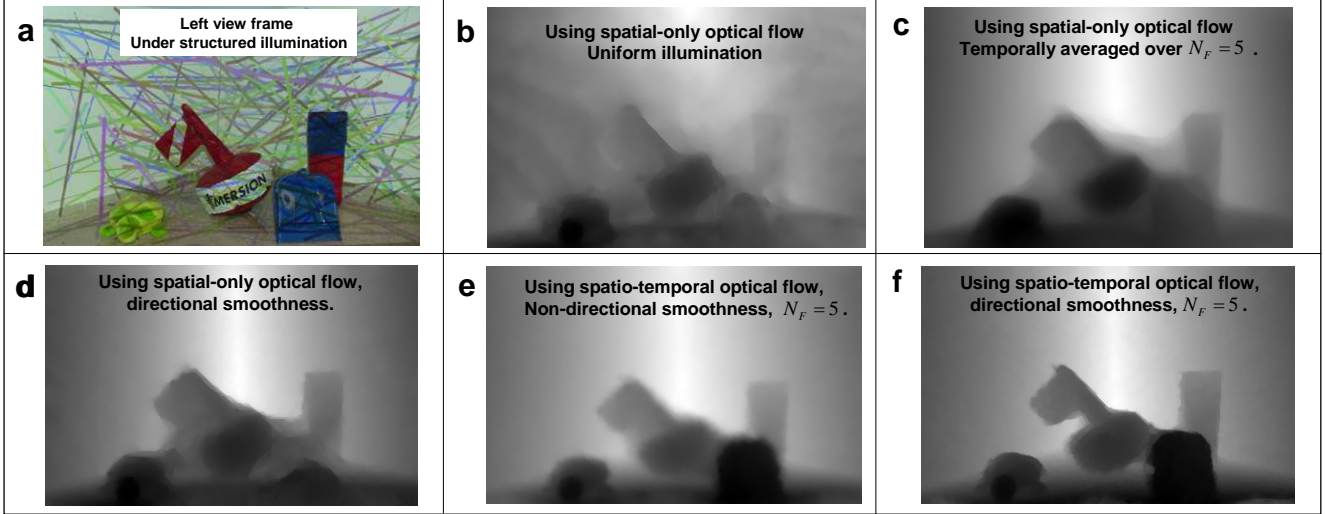


Figure 2. [a] Left viewpoint at one instance in a structured light experiment. The following are inverse disparity maps (IDMs) based on the optical flow formulation, using different processing methods. [b] Spatial-only flow, using uniform illumination. [c] Temporal average ($N_F = 5$) of IDMs, each based on spatial-only flow (using structured light) and a directional smoothness term. [d] Spatial-only flow, using structured light and directional smoothness. [e] Spatio-temporal flow, structured light, $N_F = 5$, L_1 non-directional smoothness term. [f] Spatio-temporal flow, structured light, $N_F = 5$, directional smoothness term.

approximation of the data and smoothness terms, we used the values $\varepsilon_D = 7$ and $\varepsilon_S = 0.1$, respectively.

The update rate parameter is $n_{\text{update}} = 30$. The total number of iterations at each resolution level is $N_{\text{iter}} = 200$. The coarsest resolution is 6×6 . The inter-scale shrink factor (used to create the coarse-to-fine representation), γ , is different between the axes: it is the smallest value above 0.7 that fits the coarsest level size. The number of scales is determined according to the shrink factor and coarsest resolution. Finally, unless specifically mentioned otherwise, the directional smoothness term was used in all the experiments.

4.2. Structured Light

Similar to Refs. [9, 40], the algorithm can be used in artificial spatio-temporal varying illumination. The main benefit of our method compared to Refs. [9, 40] is regularization through a directional L_1 smoothness term. This yields estimation in shadows and occluded areas and enables a shorter acquisition time. In a structured light experiment, the setup is shown in Fig. 1.b. This scene contains large areas which lack natural texture. The projected light template consisted of random lines of different lengths and orientations.

Fig. 2 displays the main results, in the form of inverse disparity maps (IDMs). Note the depth edges. Clearly, the light pattern significantly aids in finding stereo correspondences, as can be noticed in Figs. 2b and Fig. 2d. Fig. 2c shows that a temporally averaged disparity map of spatial-only optical flow blurs depth details. The significant improvement of a directional smoothness term compared to

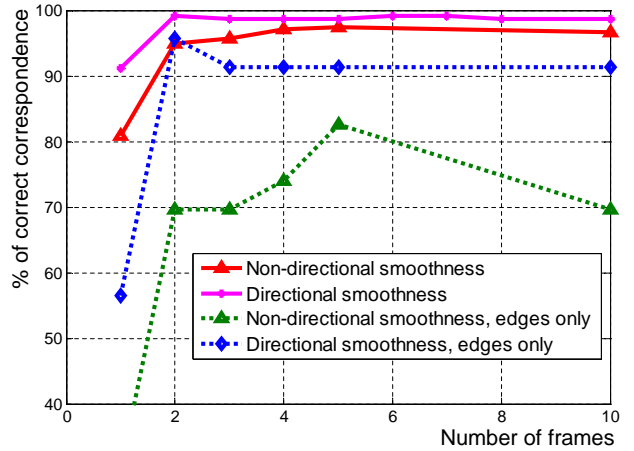


Figure 3. The rate of successful match as function of N_F in a structured light experiment, using directional and non-directional smoothness.

non-directional L_1 smoothness is clearly noticed in Figs. 2e and 2f. Fig. 3 presents quantitative results of correct correspondence as function of N_F using directional and non-directional smoothness terms.

4.3. A Dynamic Underwater Experiment

As mentioned in Sec. 1, we believe the method has the ability to handle a certain level of camera and scene motions. This is due to a relatively small N_F . We thus performed an underwater dynamic experiment. Our setup consists of two Canon HV-30 High-definition video cameras in

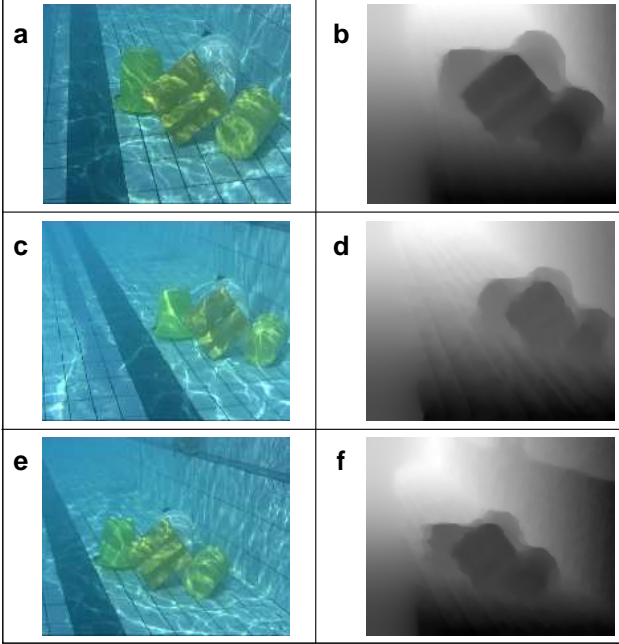


Figure 4. [a], [c], [e] Three frames of the left view, in a dynamic pool experiment. [b], [d], [f] The respective IDMs, based on spatio-temporal optical flow, using $N_F = 3$.

water-proof housings, mounted on a stereo rig. The stereo rig was handled manually during swimming, without any use of a tripod. This is contrary to the stable system used in Ref. [31]. To show the dynamics of the scene, several frames and their respective inverse disparity results are presented in Fig. 4.

If a scene is too dynamic, then the assumption of equal disparity maps between consecutive frames is no longer valid. In this case, artifacts may appear in the resulting disparity map, as can be subtly noticed on the pool floor in Fig. 4d. Such a strong camera motion may require image stabilization, prior to estimating the disparity map.

4.4. Underwater Experiments: Static Objects

In order to obtain quantitative results, we use online data [26] of 100 ground-truth points and correlation results from a swimming pool experiment [31] in a static scene. Fig. 5 compares the rate of correct correspondence results using temporal correlation, spatiotemporal correlation and our method, as a function of N_F . Our method achieves a significant recovery in $N_F \in [1, 5]$.

The result of an oceanic experiment is presented in Fig. 6. The visibility was poorer than the pool and the temporal variations of flicker were dimmer. Our method, using $N_F = 3$, achieved significantly better results than spatiotemporal correlation results, based on $N_F = 66$. Moreover, variational methods “fill-in” uncertainty regions, such as shadows or occlusions. This suppresses outliers and re-

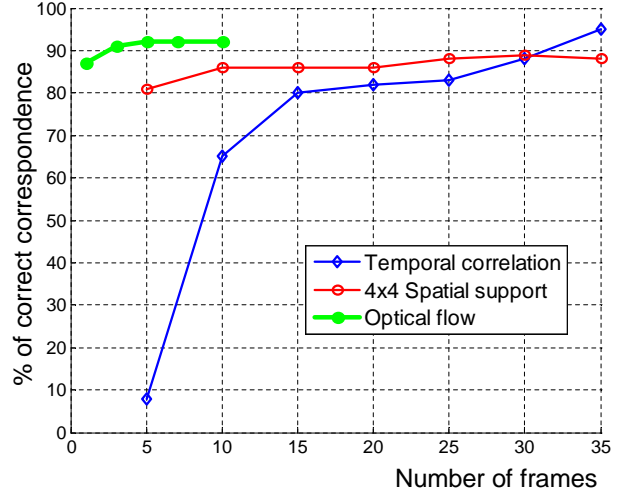


Figure 5. The rate of successful match as function of N_F in a swimming pool experiment, where the scene is static. The runs use either spatiotemporal correlation or a spatiotemporal optical flow formulation.

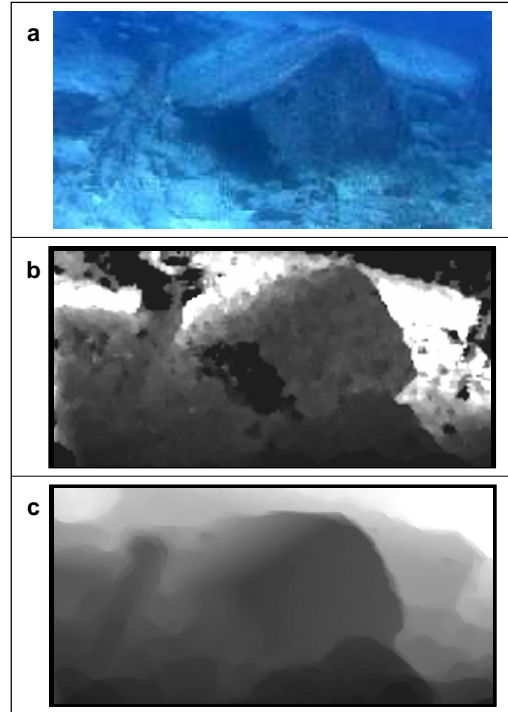


Figure 6. [a] A left frame from an experiment in a marine archaeological site (Caesarea) [31]. [b] IDM created using spatiotemporal correlation, $N_F = 66$. [c] IDM created using spatiotemporal optical flow formalism, $N_F = 3$.

gions of incorrect correspondences, which often appear in uncertain regions in correlation-based methods.

This may also be a disadvantage. Filling-in of uncertain regions may bias the results. An example is shown in

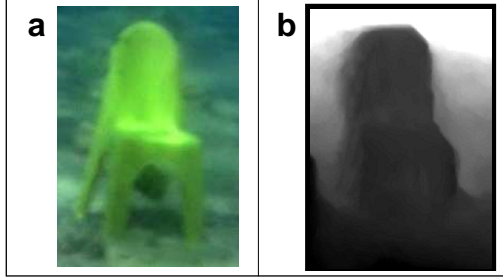


Figure 7. Example of an erroneous “fill-in” in a shadow. [a] Left viewpoint of a chair underwater. [b] IDM created using a spatio-temporal optical flow formalism, $N_F = 5$. The shadowed region under the chair is physically farther from the camera than the foreground chair parts, but the estimated depth map does not sense this concavity.

Fig. 7. There, the resulting disparity map interpolates incorrectly the disparity at a region under a chair. This region is shadowed. So, the temporal variations in it are negligible. Only regularization supports the estimation there, relying on the surrounding foreground parts of the chair. This biases the result such that the concavity under the chair is not sensed.

4.5. Numerical Scheme Validation

We wish to compare and validate the numerical scheme to existing optical flow methods. We used the standard Yosemite [3] sequence without clouds. We ran our method using the same parameters described on Sec. 4.1, on two consecutive frames. This yields an averaged angular error (AAE) of $AAE = 1.66^\circ$, compared to $AAE = 1.59^\circ$ from the 2D case in Ref. [6]. Therefore, our results are comparable to Ref. [6].

5. Computational Load

The computational complexity of the algorithm is $\mathcal{O}(WHN_F N_{\text{iter}}/[1 - \gamma^2])$, where W and H are the resolution width and height, respectively. To evaluate the runtime, a benchmark test was conducted on the oceanic experiment shown in Fig. 6. The parameters used in the benchmark are the same as in Sec. 4.1. The frame resolution of the oceanic experiment is 360×480 . The benchmark was conducted on Intel Core 2 Duo[®], 2.8GHz. The algorithm was implemented in C++. Under these conditions, the runtime was $0.66N_F$ seconds, per instantaneous disparity map.

6. Conclusions

In this work, a variational formulation for finding dense stereo correspondence using several pairs of frames is described. The formulation exploits spatio-temporal varying illumination, which occurs using structured light or naturally underwater. Moreover, using a relatively small N_F

and the inherent robustness of the variational formulation, a certain level of camera and scene motions is handled.

No epi-polar constraints were considered in this work.⁵ This fact emphasizes the robustness of the method. The formulation can be adjusted to account for epi-polar constraints, if known, as shown in [29]. Moreover, the resulting dense disparity map of the method can aid in finding the epi-polar geometry of the stereo rig, as shown in [19].

Optimization-based methods such as graph cuts [5], belief propagation [30] and stereo based on optical flow [29] have not been adapted yet to a multi-frame formulation. Out of the range of known minimization formulations, we chose to use a variational method based on optical flow. This is because of its similar complexity to uncalibrated stereo, where the search for disparity is not bounded to epi-polar lines. Nevertheless, we believe that there is room for exploring and adapting the other minimization formulations to multi-frame stereo.

Acknowledgements

We thank Tali Treibitz and Izchak Yogev for their help in conducting the underwater experiments, and Ayellet Tal and Pietro Perona for useful discussions. Yoav Schechner is a Landau Fellow - supported by the Taub Foundation. The work was supported in part by Department of the Navy Grant N62909-10-1-4056 issued by the Office of Naval Research Global. It was also partly supported by the Israel Science Foundation (Grant 1031/08), and involved work in the Ollendorff Minerva Center. Minerva is funded through the BMBF.

References

- [1] P. A. Beardsley, J. M. Brady, and D. W. Murray. Prediction of stereo disparity using optical flow. In *Proc. BMVC*, pages 259–264, 1990.
- [2] M. Black and P. Anandan. Robust dynamic motion estimation over time. In *Proc. IEEE CVPR*, pages 292–302, 1991.
- [3] M. J. Black. Yosemite sequence without clouds. <http://www.cs.brown.edu/people/black/images.html>
- [4] T. Boulton. DOVE: Dolphin omni-directional video equipment. In *Proc. IASTED Int. Conf. Robotics and Autom.*, pages 214–220, 2000.
- [5] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. PAMI*, 23:1222–1239, 2001.
- [6] T. Brox, A. Bruhn, N. Papenberger, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *Proc. ECCV*, pages 25–36, 2004.

⁵The water interface introduces a non-single viewpoint geometry [34, 35] in each camera. This makes underwater epi-polar geometry more complex [8].

- [7] 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.*, pages 111–116, 2000.
- [8] V. Chari and P. Sturm. Multi-view geometry of the refractive plane. In *Proc. BMVC*, 2009.
- [9] J. Davis, D. Nehab, R. Ramamoorthi, and S. Rusinkiewicz. Spacetime stereo: A unifying framework for depth from triangulation. *IEEE Trans. PAMI*, 27:296–302, 2005.
- [10] M. Finckh, H. Dammertz, and H. Lensch. Geometry construction from caustic images. In *Proc. ECCV*, pages 464–477, 2010.
- [11] G. L. Foresti. Visual inspection of sea bottom structures by an autonomous underwater vehicle. *IEEE Trans. Syst. Man and Cyber*, 31:691–705, 2001.
- [12] 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.
- [13] H. W. Haussecker and D. J. Fleet. Computing optical flow with physical models of brightness variation. *IEEE Trans. PAMI*, 23:661–673, 2001.
- [14] B. K. P. Horn and B. G. Schunck. Determining optical flow. *AI*, 17:185–203, 1981.
- [15] S. Hrabar and G. S. Sukhatme. Combined optic-flow and stereo-based navigation of urban canyons for a uav. In *Proc. IEEE/RSJ IROS*, pages 2–6, 2005.
- [16] H. Jeffreys and B. S. Jeffreys. *Methods of Mathematical Physics*, pages 305–306. Cambridge University Press, 3rd edition, 1988.
- [17] 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.*, 42:52–67, 2008.
- [18] J. M. Lavest, F. Guichard, and C. Rousseau. Multiview reconstruction combining underwater and air sensors. In *Proc. IEEE ICIP*, volume 3, pages 813–816, 2002.
- [19] M. Mainberger, A. Bruhn, and J. Weickert. Is dense optic flow useful to compute the fundamental matrix? In *Proc. ICAR*, 2008.
- [20] W. N. McFarland and E. R. Loew. Wave produced changes in underwater light and their relations to vision. *Env. Biol. Fish.*, pages 173–184, 1983.
- [21] H. H. Nagel. Extending the 'oriented smoothness constraint' into the temporal domain and the estimation of derivatives of optical flow. In *Proc. ECCV*, pages 139–148. Springer, Berlin, 1990.
- [22] S. Negahdaripour. Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis. *IEEE Trans. PAMI*, 20:961–979, 1998.
- [23] S. Negahdaripour and P. Firoozfam. An ROV stereovision system for ship-hull inspection. *IEEE JOE.*, 31:551–564, 2006.
- [24] T. Nir and N. Karpel. Example based learning of image stitching for an omni directional camera using a variational optical flow methodology. In *Proc. SPIE Int. Soc. of Opt. Eng.*, volume 7000, pages 700024–700024–11, 2008.
- [25] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47:7–42, 2002.
- [26] Y. Y. Schechner, Stereo from flicker. [Online data]. <http://webee.technion.ac.il/~yoav/research>
Follow the link to *Underwater CauStereo*.
- [27] Y. Y. Schechner and N. Karpel. Recovery of underwater visibility and structure by polarization analysis. *IEEE J. Oceanic Eng.* 30:570-587, 2005.
- [28] Y. Y. Schechner and N. Karpel. Attenuating natural flicker patterns. In *Proc. MTS/IEEE Oceans*, pages 1262–1268, 2004.
- [29] N. Slesareva, A. Bruhn, and J. Weickert. Optic flow goes stereo: A variational method for estimating discontinuity-preserving dense disparity maps. In *Pattern Recognition*, pages 33–40, 2005.
- [30] J. Sun, N. Zheng, and H. Shum. Stereo matching using belief propagation. *IEEE Trans. PAMI*, 25:787 – 800, 2003.
- [31] Y. Swirski, Y. Y. Schechner, B. Herzberg, and S. Negahdaripour. Stereo from flickering caustics. In *Proc. IEEE ICCV*, 2009.
- [32] Y. Swirski, Y. Y. Schechner, B. Herzberg, and S. Negahdaripour. Underwater stereo using natural flickering illumination. In *Proc. IEEE/MTS Oceans*, 2010.
- [33] Y. Swirski, Y. Y. Schechner, B. Herzberg, and S. Negahdaripour. CauStereo: Range from light in nature. *Applied Optics*, to appear, 2012.
- [34] T. Treibitz, Y. Y. Schechner, and H. Singh. Flat refractive geometry. In *Proc. IEEE CVPR*, 2008.
- [35] T. Treibitz, Y. Y. Schechner, C. Kunz and H. Singh, Flat refractive geometry. *IEEE Trans. PAMI*, to be published, 2011.
- [36] Y. Tian and S. G. Narasimhan. Seeing through water: image restoration using model-based tracking. In *Proc. IEEE ICCV*, 2009.
- [37] L. Valgaerts, A. Bruhn, and J. Weickert. A variational model for the joint recovery of the fundamental matrix and the optical flow. In *DAGM*, pages 314–324, 2008.
- [38] R. E. Walker. *Marine Light Field Statistics*, chapter 10. John Wiley, New York, 1994.
- [39] J. Weickert and C. Schnörr. Variational optic flow computation with a spatio-temporal smoothness constraint. *J. Mathematical Imaging and Vision*, 14:245–255, 2001.
- [40] L. Zhang, B. Curless, and S. Seitz. Spacetime stereo: shape recovery for dynamic scenes. In *Proc. IEEE CVPR*, pages 367–374, 2003.