Proc. IEEE Int. Conference on Computational Photography (ICCP) 2013

3Deflicker from Motion

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

Abstract

Spatio-temporal irradiance variations are created by some structured light setups. They also occur naturally underwater, where they are termed flicker. *Underwater, visibility is also affected by water scattering. Methods for overcoming or exploiting flicker or scatter exist, when the imaging geometry is static or quasi-static. This work removes the need for quasi-static scene-object geometry under flickering illumination. A scene is observed from a free moving platform that carries standard frame-rate stereo cameras. The 3D scene structure is illumination invariant. Thus, as a reference for motion estimation, we use projections of* stereoscopic range maps*, rather than object radiance. Consequently, each object point can be tracked and then filtered in time, yielding deflickered videos. Moreover, since objects are viewed from different distances as the stereo rig moves, scattering effects on the images are modulated. This modulation, the recovered camera poses, 3D structure and deflickered images yield inversion of scattering and recovery of the water attenuation coefficient. Thus, coupled difficult problems are solved in a single framework. This is demonstrated in underwater field experiments and in a lab.*

1. Introduction

Dynamic spatiotemporal scene irradiance is used in structured-light setups [5, 12, 15, 31, 55]. It also exists naturally underwater [17, 51], where the light pattern is uncontrolled and unknown. This pattern is often referred to as *sunlight flicker* [11] or *caustic networks* [29]. An example is shown in Fig. 1. The pattern is created by refraction of sunlight through the dynamic wavy air-water[1] surface [6, 8, 9, 48]. Of course, flicker strongly affects underwater vision. This computer vision domain is essential for man-made systems handling oceanic engineering tasks [22], archaeology [20] and mapping. This domain is also important for understanding biological vision of marine animals [30].

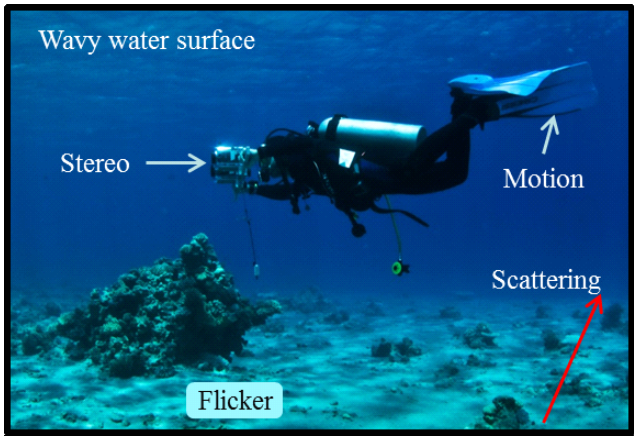

Figure 1. An underwater scene has natural spatiotemporally varying unknown illumination (flicker). The scene is also affected by scattering, which accumulates through the unknown object distance and attenuation parameters. The scene is captured by a stereo rig that has unknown free motion. We seek to recover motion and scene structure, to deflicker and descatter the scene.

Inconsistent spatiotemporal irradiance variations confuse visual analysis and understanding based on object reflectance. However, these variations are very helpful for three dimensional (3D) shape reconstruction, whether by structured [31, 36, 39] or unstructured [5, 24] light setups. Unstructured temporal flicker easily establishes correspondence in stereoscopic videos [46, 47]. Moreover, as each object point is measured multiple times in a sequence, flickering per object point can be filtered-out over time, creating de-flickered images [11, 40].

Nevertheless, such setups have so far assumed that during image-sequence capture, the camera-object geometry is static or quasi-static [55]. The motion has been small enough to be tolerated in a very small number of frames. These methods trade-off some spatial resolution, to enable handling temporal geometry changes. Alternatively, fast camera-projector setups were used for structured light [23].

[1] Stereoscopic imaging [2] and motion analysis [1] through a wavy air-water interface are studied as well.

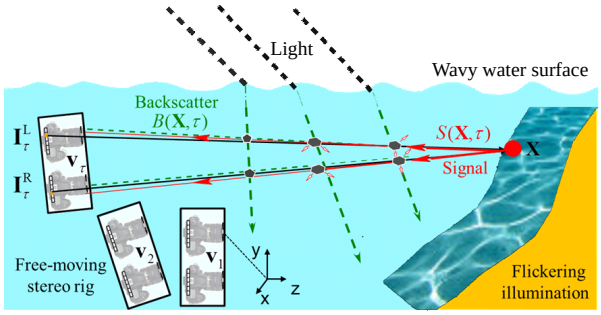

Figure 2. An underwater moving stereo rig. Images are affected by scattering, signal attenuation, dynamic illumination, rig motion and viewpoint parallax. The global 3D coordinates are aligned with the left camera at $\mathbf{v}_1$.

This work removes the need for quasi-static scene-object geometry under flickering illumination. The scene is observed from a free moving platform, using standard frame-rate cameras. The work enables structure from stereo and motion [53, 54] in dynamic illumination. The method simultaneously yields camera pose estimation [16] and dense 3D scene structure. This generalizes structure from motion (SfM) [44] to handle the challenging illumination field. Consequently, each object point can be *tracked* and *geometrically stabilized* in time. As a result, the flickering radiance readings of each point can be filtered temporally, yielding deflickered videos.

Underwater, this algorithm has an additional important benefit: it enables countering and analyzing scattering effects. Underwater scattering and absorption degrade visibility and color. For reliable compensation (descattering [49]), scattering effects should be modulated between frames. We obtain this modulation naturally: objects are viewed from different distances as the rig moves, and this varies the effect of scattering on the images. This modulation, the recovered camera poses, 3D structure and deflickered images yield inversion of scattering.[2] Thus, coupled difficult problems are solved in a single framework. This is demonstrated in underwater field experiments and in a lab.

2. Theoretical Background

2.1. Underwater Scatter

Underwater images suffer from strong scattering and signal attenuation. The images comprise an attenuated signal and light backscattered along the line of sight (LOS). This is illustrated in Fig. 2. The signal S is the object radiance that would have been measured if there was no water on the LOS. The attenuation coefficient of the water is η, while r is the distance of the object from a submerged camera. Under still natural illumination, the measured radiance [41] is

$$i = Se^{-\eta r} + B_{\infty}[1 - e^{-\eta r}] \ , \tag{1}$$

where B_{∞} is the total backscatter radiance at $r \to \infty$. The global parameters η and B_{∞} depend on the light wavelength, thus affecting color. Descattering recovers S. This requires estimations of r for each pixel, and B_{∞} and η for each color channel. Given these estimates, a descattered radiance is calculated per visible object point $\mathbf{x}$ by

$$\hat{S}(\mathbf{x}) = \frac{i(\mathbf{x}) - B_{\infty}[1 - e^{-\eta r(\mathbf{x})}]}{e^{-\eta r(\mathbf{x})}} \ . \tag{2}$$

Descattering can be assisted by stereo [32, 37] if range is estimated by stereo triangulation.

2.2. Stereo Correspondence and Triangulation

Stereo vision can yield the 3D structure of the scene. This task has two sub-problems: establishing correspondence and triangulation [13]. Stereo correspondence is well-studied in computer vision [38]. Once correspondence between two viewpoints is achieved, triangulation between every matching pair of pixels can be calculated in order to estimate the corresponding 3D object point. Metric scene reconstruction requires calibration of both intrinsic and extrinsic parameters [13, 56] of the stereo rig.

2.3. Alignment of Point Clouds

The column vector $\mathbf{x}$ expresses the coordinates of a 3D point in a global coordinate system. The $3 \times N$ matrix $\mathbf{X}$ expresses an N-point cloud in 3D. This cloud is obtained by sensing a rigid object from pose $\mathbf{v}_1$. The same object is measured from another viewpoint, $\mathbf{v}_2$, yielding another point cloud, $\mathbf{X}'$. To align the two clouds, there is a need to estimate the rigid, 6-degrees of freedom (DOF) transformation (rotation and translation) between the point clouds. This is equivalent to finding the 6-DOF transformation between $\mathbf{v}_1$ and $\mathbf{v}_2$. The DOFs can be expressed in a rotation matrix $\mathbf{R}$ and a translation vector $\mathbf{t}$, which represent the transformation between the two point clouds. Estimation can be done by minimizing a cost function:

$$[\hat{\mathbf{R}}, \hat{\mathbf{T}}] = \arg\min_{\mathbf{R},\mathbf{t}} \{\|\mathbf{X} - \mathbf{R}\tilde{\mathbf{X}}' - \mathbf{t}\|_2\} \ . \tag{3}$$

Here $\tilde{\mathbf{X}}'$ is a re-ordered version of $\mathbf{X}'$, such that corresponding points (columns) of $\tilde{\mathbf{X}}'$ and $\mathbf{X}$ are nearest neighbors.

Equation (3) can be solved using iterative closest point (ICP) algorithms [35, 43]. They iteratively match points in the two clouds, then re-assess the 6-DOF transformation to minimize Eq. (3). Fig. 3 demonstrates alignment resulting from an ICP process.[3]

[2] Descattering in open air (dehazing and defogging) is widely studied [7, 14, 19, 33].

[3] Registration of several point clouds can be done simultaneously. This process is commonly referred to as *global registration* [34].

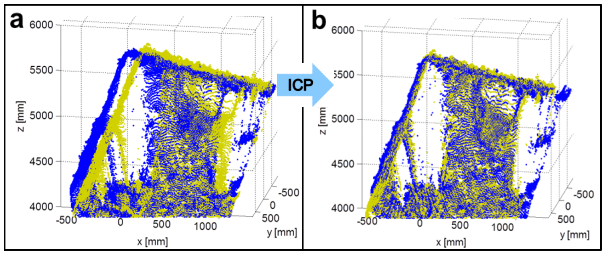

Figure 3. Alignment following 6-DOF rigid motion estimation using ICP. [a] Two point clouds (blue and green) of an object, each generated by stereoscopic triangulation from a distinct rig pose. [b] The same point clouds after rigid alignment.

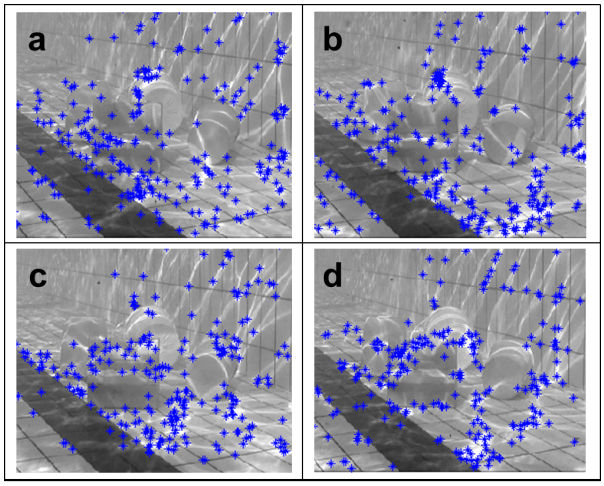

Figure 4. Consecutive frames of a static scene in a swimming pool affected by flicker. The blue asterisks are key points detected by SIFT. Most key points are inconsistent between frames, due to the strong illumination variations.

3. The Challenge of Refractive Flicker

Video stabilization algorithms usually rely on block-based search [21] or key points match [3, 26]. To illustrate the challenge in our scenarios, we analyzed an underwater video sequence in a swimming pool, where the video camera was *static*. Sample frames are shown in Fig. 4. SIFT [28] features were used to find corresponding key points between frames. Although the camera and objects were static, there are almost no static key points, due to the strong spatiotemporal lighting variations. This can mislead motion estimation. To show this, we rendered a vibrating sequence of frames from this sequence. As shown in Fig. 5, the estimated camera motion has significant errors, compared to the simulated ground-truth.

4. Principle

To estimate camera motion, we need a static object of reference. While the object radiance may strongly vary

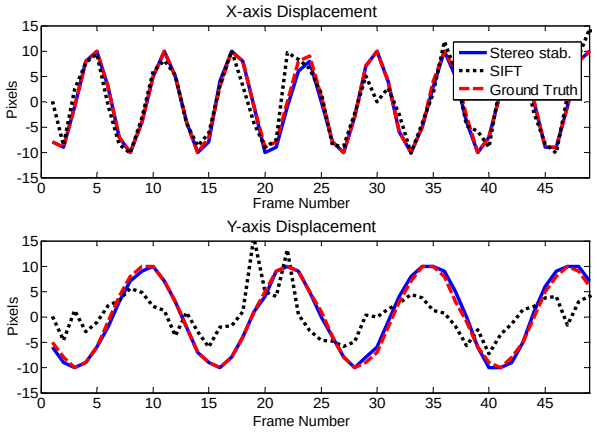

Figure 5. SIFT-based estimation (dotted line) of displacement in a vibrating video sequence is erroneous. Proper estimation is achieved if the vibrating rig has stereo cameras, and if the stereoscopic range map functions as the alignment target.

under dynamic illumination, the 3D scene structure is illumination invariant. Therefore, as a static object of reference, we use projections of *stereoscopic range maps*, rather than object radiance.[4] We thus acquire two simultaneous video sequences by a moving stereo rig, rather than a monocular video. Relying on range maps as the alignment target is robust to lighting variations, but a bit coarse. The reason is that range maps are typically less textured than the object radiance. Thus, refinement is achieved by exploiting the spatio-temporal irradiance variations, as in Refs. [45, 46, 47].

To demonstrate this principle, the setup that acquired the video referred to in Figs. 4,5 included *two video cameras* on a static rig. We used the videos to render two corresponding vibrating sequences. For each simultaneous pair of frames, a stereoscopic range map was calculated and used for video stabilization. The rendered frames of the left viewpoint are shown in Figs. 6a-c, the corresponding depth maps are shown in Figs. 6d-f and the stabilized results are shown in Figs. 6g-h. The estimated motion is consistent with the ground truth, as plotted in Fig. 5. The stabilized video sequences can then be used to attenuate the flicker [11, 40] (Fig. 6i).

5. Stereoscopic SfM in Dynamic Illumination

The method described in Sec. 4 is generalized to 6-DOF. First, per time τ, the stereo rig yields a 3D point cloud $\mathbf{X}_\tau$. ICP is applied on the point clouds $\{\mathbf{X}_\tau\}_\tau$, to estimate the 6-DOF motion of the stereo rig. Second, as in Ref. [46], the *temporal* variations of illumination enhance the accuracy of the estimation. We now detail the main elements we used.

[4] This principle is strongly related to Refs. [25, 27], where depth maps are found useful for image stabilization and intrinsic images estimation.

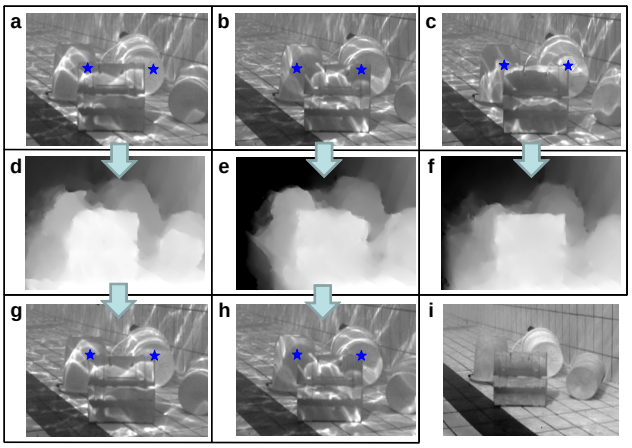

Figure 6. [a-c] Frames from the left viewpoint of a static stereoscopic video sequence in a swimming pool featuring rotational artificial movement. The blue asterisks mark the corner positions of the chest in the first frame. [d-f] The corresponding disparity maps of frames [a-c], respectively. [g-h] Frames corresponding to [a-b], respectively, stabilized to align with frame [c]. The blue asterisks positions can be used to evaluate stabilization accuracy. [i] A deflickered image rendered by temporal median on 15 stabilized frames.

Stereo using a simultaneous pair of frames: A stereo rig is calibrated [56]. Afterwards, during scene acquisition, the rig is at an unknown viewpoint pose $\mathbf{v}_\tau$, where τ is the discrete temporal index. The image in the left camera then is $\mathbf{I}^{\mathrm{L}}_\tau$. Simultaneously, the image in the right camera is $\mathbf{I}^{\mathrm{R}}_\tau$. The illumination field is the same for both simultaneous views. Hence, stereo correspondence can be established using any algorithm [38, 47] for calibrated stereo correspondence. Given the calibration and correspondence field, triangulation yields a 3D point-cloud $\mathbf{X}_\tau$. Each point in the cloud is an estimated sample of the object's 3D structure.

Rig pose estimation using ICP: Let there be a prior point cloud $\mathbf{X}^{\mathtt{model}}$, to which $\mathbf{X}_\tau$ needs to be aligned. The prior cloud $\mathbf{X}^{\mathtt{model}}$ is first obtained by an initialization step, and afterwards updated with incoming data, as we describe below. ICP is applied on $\mathbf{X}_\tau$ and $\mathbf{X}^{\mathtt{model}}$, to recover the pose (translation and rotation) $\mathbf{v}_\tau \equiv \{\mathbf{R}_\tau, \mathbf{t}_\tau\}$ in the global coordinate system.[5]:

$$[\hat{\mathbf{R}}_\tau, \hat{\mathbf{t}}_\tau] = \arg\min_{\mathbf{R},\mathbf{t}} \{\|\mathbf{X}^{\mathtt{model}} - \mathbf{R}\tilde{\mathbf{X}}_\tau - \mathbf{t}\|_2\} \ . \tag{4}$$

Warping the video frames to pose $\mathbf{v}_\tau$: This step seeks to stabilize the left and right videos, as if they were acquired by a static rig whose pose is $\mathbf{v}_\tau$. Here is the process we employed. The process assumes that $\mathbf{X}^{\mathtt{model}}$ represents the object structure more reliably than any particular cloud $\mathbf{X}_{\tau'}$ of any time τ'. However, $\mathbf{X}^{\mathtt{model}}$ lacks radiance texture. Thus, $\mathbf{X}^{\mathtt{model}}$ is texture-mapped, by computationally back-projecting the left image $\mathbf{I}^{\mathrm{L}}_{\tau'}$ onto $\mathbf{X}^{\mathtt{model}}$. This back-projection is possible since $\mathbf{v}_{\tau'}$ has already been estimated. The textured-map cloud can be represented as a matrix

$$\mathbf{C}^{\mathrm{L}}_{\tau'} \equiv \begin{bmatrix} \tilde{\mathbf{X}}_{\tau'} \\ \tilde{\mathbf{I}}^{\mathrm{L}}_{\tau'} \end{bmatrix} = \mathtt{BackProject}^{\mathrm{L}}(\mathbf{I}^{\mathrm{L}}_{\tau'}, \mathbf{X}^{\mathtt{model}} | \mathbf{v}_{\tau'}) \ . \tag{5}$$

Recall that $\tilde{\mathbf{X}}_{\tau'}$ is a re-ordered version of $\mathbf{X}_{\tau'}$, such that corresponding points of $\tilde{\mathbf{X}}_{\tau'}$ and $\mathbf{X}^{\mathtt{model}}$ are nearest neighbors. Here, $\tilde{\mathbf{I}}^{\mathrm{L}}_{\tau'}$ is a $1 \times M$ row-vector reordering of the pixels in $\mathbf{I}^{\mathrm{L}}_{\tau'}$ where each pixel corresponds a column in $\tilde{\mathbf{X}}_{\tau'}$.

Then, $\mathbf{C}^{\mathrm{L}}_{\tau'}$ is computationally projected to the left camera, as if the rig is at $\mathbf{v}_\tau$. This yields the warped left image

$$\mathbf{I}^{\mathrm{L},\mathtt{warp}}_{\tau' \to \tau} = \mathtt{Project}^{\mathrm{L}}(\mathbf{C}^{\mathrm{L}}_{\tau'} | \mathbf{v}_\tau) \ . \tag{6}$$

Analogously, the right image $\mathbf{I}^{\mathrm{R}}_{\tau'}$ is back-projected to yield a cloud $\mathbf{C}^{\mathrm{R}}_{\tau'}$, which is then projected to yield a warped right-image, $\mathbf{I}^{\mathrm{R},\mathtt{warp}}_{\tau' \to \tau}$. This process is done $\forall \tau' \neq \tau$. The output is a geometrically stabilized left-cam video $\{\mathbf{I}^{\mathrm{L},\mathtt{warp}}_{\tau' \to \tau}\}_{\tau'}$ and a stabilized right-cam video $\{\mathbf{I}^{\mathrm{R},\mathtt{warp}}_{\tau' \to \tau}\}_{\tau'}$.

CauStereo: Refs. [46, 47] showed that in flickering illumination created by caustic networks, spatio-temporal radiance variations are very useful to obtain accurate correspondence. Refs. [46, 47] used a static stereo rig, where the left and right videos are geometrically stable. To exploit this principle in our dynamic cases, we use the digitally stabilized videos $\{\mathbf{I}^{\mathrm{L},\mathtt{warp}}_{\tau' \to \tau}\}_{\tau'}$ and $\{\mathbf{I}^{\mathrm{R},\mathtt{warp}}_{\tau' \to \tau}\}_{\tau'}$, described above, as data. On this warped data, we ran variational CauStereo [47] (a form of space-time stereo). The result is a 3D point cloud $\mathbf{X}^{\mathtt{spacetime}}_\tau$. It has a finer quality than $\mathbf{X}_\tau$, since $\mathbf{X}^{\mathtt{spacetime}}_\tau$ is triangulated based on several frame-pairs, in each of which the irradiance texture is distinct, while $\mathbf{X}_\tau$ relies only on a single frame-pair.

Updating the model cloud: The fine 3D point-cloud $\mathbf{X}^{\mathtt{spacetime}}_\tau$ can now update $\mathbf{X}^{\mathtt{model}}$. First, relying on Eq. (4), $\mathbf{X}^{\mathtt{spacetime}}_\tau$ is aligned to $\mathbf{X}^{\mathtt{model}}$, by

$$\mathbf{X}^{\mathtt{spacetime}}_{\tau \to \mathtt{model}} = \hat{\mathbf{R}}^{-1}_\tau (\mathbf{X}^{\mathtt{spacetime}}_\tau - \hat{\mathbf{t}}_\tau) \ . \tag{7}$$

Next, $\mathbf{X}^{\mathtt{spacetime}}_{\tau \to \mathtt{model}}$ is merged with $\mathbf{X}^{\mathtt{model}}$ into an updated model. This is simply done by accumulating all points into a single, large cloud.

Initializing the model cloud: As written above, $\mathbf{X}^{\mathtt{model}}$ is initialized somehow. We initialize it simply by using $\mathbf{X}_{\tau=1}$, which corresponds to the first stereo frame-pair. Without loss of generality, we align the global 3D coordinates $\mathbf{x} = [x, y, z]^T$ with the left camera, when the rig is at $\mathbf{v}_{\tau=1}$: the x and y axes are parallel to the axes of $\mathbf{I}^{\mathrm{L}}_{\tau=1}$, while the z axis is the optical axis of this camera then.

[5] For efficient estimation of $\mathbf{v}_\tau$, it is initialized by $\mathbf{v}_{\tau-1}$, since viewpoint change between consecutive frames is typically small.

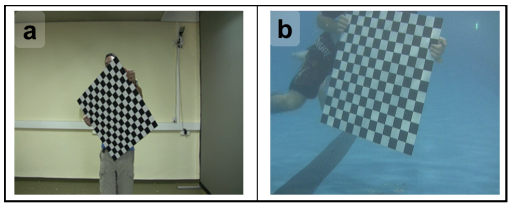

Figure 7. Sample calibration frames of the lab [a] and the underwater [b] stereo rigs.

SfM Example: Lab Experiment

Consider the first experiment conducted in the lab. Our setup consists of two Canon HV-30 camcorders. To mutually synchronize the sequences, we shined brief light flashes into the running camcorders before and after each experiment. These flashes were later detected in postprocessing and used to temporally align the videos. Camera calibration was done using a checkerboard pattern, captured by both cameras simultaneously in different poses. Details of the calibration process using a flat checkerboard can be found in Ref. [56]. An example of a calibration frame is shown in Fig 7a. Moreover, radiometric calibration was conducted in the lab, to enable oceanic descattering.

A spatiotemporal varying illumination pattern was projected on a scene, while a stereo rig moved rigidly around the scene. Two frames acquired by the moving left camera are shown in Figs. 8a,b. he entire stereoscopic sequence was used to estimate the rig trajectory and a dense structure of the scene, by the algorithm described in Sec. 5. The results are shown in Figs. 8c,d. The readers are referred to the supplementary videos [42] for better impression of the results.

6. Deflickered Images and Video

Section 5 describes how the videos can be geometrically stabilized digitally. A geometrically stabilized video $\{\mathbf{I}^{\mathrm{L,warp}}_{\tau'\to\tau}\}_{\tau'}$ can then be used to attenuate the flicker in a rendered image. For example, a deflickered image can be estimated simply by applying temporal median [11] on N_{F} stabilized frames,

$$\mathbf{O}^{\mathrm{L}}_{\tau} = \mathtt{Median}_{\tau'}\{\mathbf{I}^{\mathrm{L,warp}}_{\tau'\to\tau}\} \quad . \tag{8}$$

Eq. (8) approximates the object appearance under still water (constant lighting), viewed through the left camera when the rig is at $\mathbf{v}_{\tau}$. Repeating this process $\forall\tau$ yields a rendered video sequence $\{\mathbf{O}^{\mathrm{L}}_{\tau}\}_{\tau}$, taken as if the camera moves along its original trajectory, but the flicker pattern is eliminated.

SfM and Deflicker Example: Pool Experiment

We conducted an underwater experiment in a swimming pool. To capture images underwater, each camera was mounted inside an Ikelite waterproof housing. The underwater experimental setup appears in Fig. 1.

Cameras calibration was done using a checkerboard underwater (Fig. 7b). In a generalized case of underwater imaging through a refractive flat port, the assumption of a single viewpoint (SVP) for each camera may be invalid [4, 18, 50]. Therefore, a more complex calibration

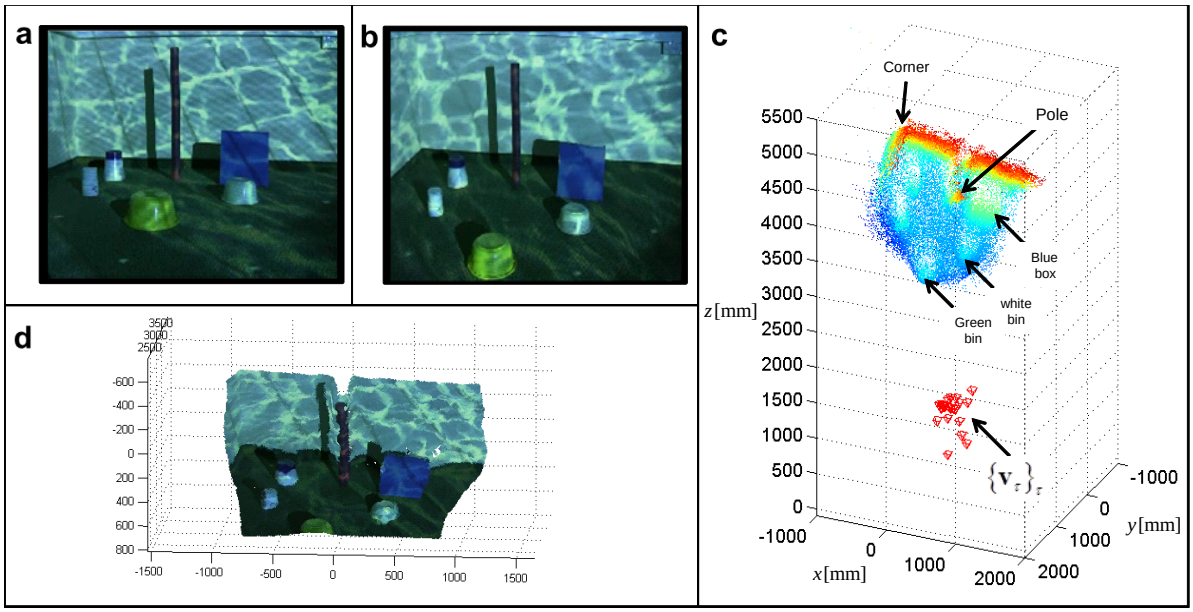

Figure 8. An indoor experiment. [a-b] Two frames from a video sequence from the left viewpoint. [c] Triangulation is used to estimate the stereo rig motion and a dense structure. [d] Texture mapping of the estimated structure using [a].

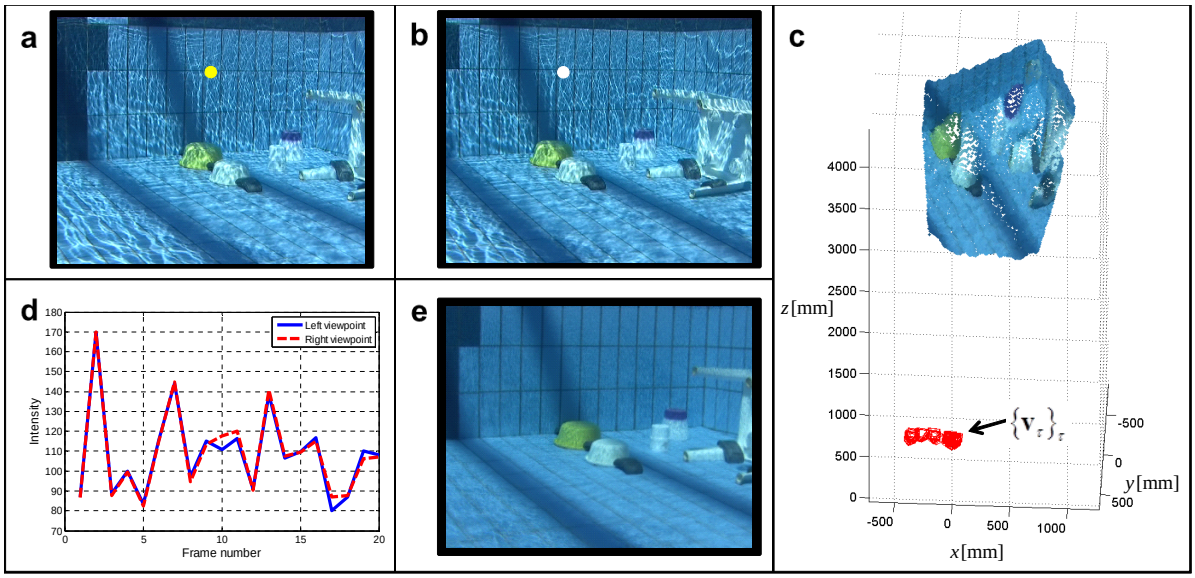

Figure 9. An underwater experiment conducted in a swimming pool. Frames from the left [a] and right [b] viewpoints from the video sequences. [c] Triangulation is used to estimate the stereo rig motion and a dense structure. [d] Temporal intensity changes of the same object point captured in the two viewpoints. The corresponding object point is marked in circles in [a] and [b]. [e] A deflickered image rendered using temporal median on a stabilized video sequence.

algorithm should be considered [10, 52]. Fortunately, in our underwater system, we found that an SVP assumption yields negligible errors and that the common model for SVP camera calibration can be used. This may be because in this system, the distance between the camera exit-pupil and the port is very small, relative to the typical scene ranges.

In the pool, the cameras were rigidly connected by a $22.5cm$ baseline. The rig was hand-held, while undergoing a free trajectory. Figs. 9a,b present two frames of the video simultaneously taken by the left and right cameras, respectively. As explained above, the method renders deflickered images. This is demonstrated in Fig. 9e. For a specific object point, temporal intensity plots measured by the two camera are extracted from $\{\mathbf{I}^{\mathtt{L,warp}}_{\tau'\to\tau}\}_{\tau'}$ and $\{\mathbf{I}^{\mathtt{R,warp}}_{\tau'\to\tau}\}_{\tau'}$. Fig. 9d compares the plots.

7. Descattering

To descatter based on Eqs. (1,2), first there is a need to assess several variables. For viewpoint $\mathbf{v}_\tau$, per visible object point $\mathbf{x}$, we need the image radiance $i(\mathbf{x})$ under flat-water. We also need the range $r_\tau(\mathbf{x})$. Moreover, estimates are needed for the global parameters B_∞ and η. The range $r_\tau(\mathbf{x})$ is easily derived $\forall \mathbf{x} \in \mathbf{X}^{\mathtt{model}}$, following the process described is Sec. 5. We use the deflickered value $O^{\mathtt{L}}_\tau(\mathbf{x})$ as an estimate for flat-water radiance $i(\mathbf{x})$, as seen from $\mathbf{v}_\tau$.

Estimation of B_∞ is based on sampling pixels [41] for which $r_\tau(\mathbf{x}) \to \infty$. However, in contrast to Ref. [41], here these pixels are perturbed by flicker. Thus, also B_∞ has to be deflickered. This is done by the same method described in Sec. 6. Specifically, if temporal median (Eq. 8) is used to deflicker, then B_∞ as seen from $\mathbf{v}_\tau$ is estimated by

$$\hat{B}_{\infty,\tau} = \mathtt{Median}_{\tau'}\{I^{\mathtt{L,warp}}_{\tau'\to\tau}(\mathbf{x})\} \quad | \; r_\tau(\mathbf{x}) \to \infty \;\; . \tag{9}$$

Recall that η needs to be estimated. Let us use a set $\tilde{\mathbf{X}}$ of 3D points. The unknown signals corresponding to these points form an array $\tilde{\mathbf{S}}$. Define a cost function

$$E_\eta(\tilde{\mathbf{S}}) = \sum_{\tau,\mathbf{x}\in\tilde{\mathbf{X}}} \mid O^{\mathtt{L}}_\tau(\mathbf{x}) + [\hat{B}_{\infty,\tau} - S(\mathbf{x})]e^{-\eta r_\tau(\mathbf{x})} - \hat{B}_{\infty,\tau} \mid^2 \tag{10}$$

Suppose for the moment that η is known. Then, Eq. (10) is minimized by nulling the gradient of Eq. (10) with respect to $\tilde{\mathbf{S}}$. This occurs at

$$\hat{S}_\eta(\mathbf{x}) = \frac{\sum_\tau \{O^{\mathtt{L}}_\tau(\mathbf{x}) + \hat{B}_{\infty,\tau}[e^{-\eta r_\tau(\mathbf{x})} - 1]\} e^{-\eta r_\tau(\mathbf{x})}}{\sum_\tau e^{-2\eta r_\tau(\mathbf{x})}} \;\; . \tag{11}$$

Substituting Eq. (11) in Eq. (10) results in a cost $E_\eta \equiv E(\hat{\mathbf{S}}_\eta)$. Now, let η be variable. Then, we are left with a one-dimensional minimization problem:

$$\hat{\eta} = \arg\min_\eta E_\eta \;\; . \tag{12}$$

The optimal result is easily searched. Using the parameters $\hat{\eta}$ and $\hat{B}_{\infty,\tau}$ in Eq. (11) yields descattered images $\forall \mathbf{x}, \tau$.

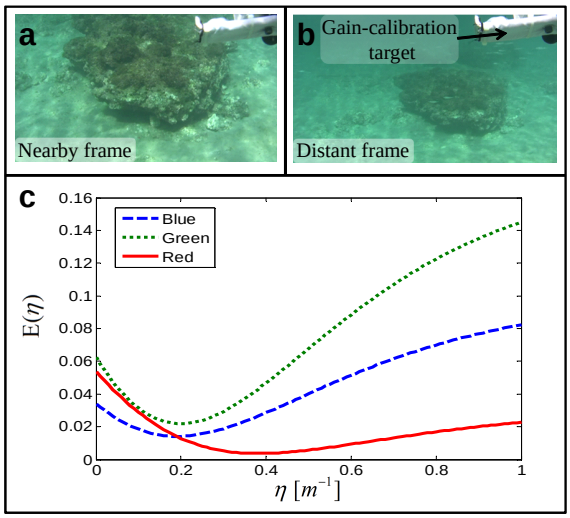

Figure 10. Two left viewpoint frames captured near [a] and far [b] from a submerged rock. The different distances from the objects modulate scattering, leading to estimation of η. A gain-calibration target was mounted on the rig. Its shadowed part was used to monitor the cameras gain changes over time. [c] Estimation of the attenuation coefficient (per color). Optimal η values are the minima in the plots.

Oceanic Experiment

We conducted an experiment in the sea, in Dor beach, Israel. The oceanic experimental setup appears in Fig. 1. Figs. 10a,b present two frames of the same object, from different ranges. As mentioned in Sec. 5, the cameras had been radiometrically calibrated. However, our camcorders have auto-gain, which affects radiometric reconstruction. Therefore, we mounted a gain-calibration target on the rig. The target is seen in the corner of the frames (Figs. 10a,b). Its shadowed part indicates the gain changes over time. Corresponding object points in different frames yield the required data for η estimation. Estimation results for the red, green and blue color channels appear in Fig. 10c. Minimization results are consistent with the average of statistical values for coastal waters [51]. The estimated parameters were then used for descattering in the same environment.

Results from the main oceanic experiment are presented in Fig. 11. The conditions in the ocean are more challenging than in a pool, as poor visibility due to scatter compounds flicker. Nevertheless, structure and motion are estimated, as shown in Fig. 11d. A deflickered image $\mathbf{O}^{\mathrm{L}}_{\tau}$ which corresponds to Fig. 11a appears in Fig. 11b. The descattered result is shown in Fig. 11c.

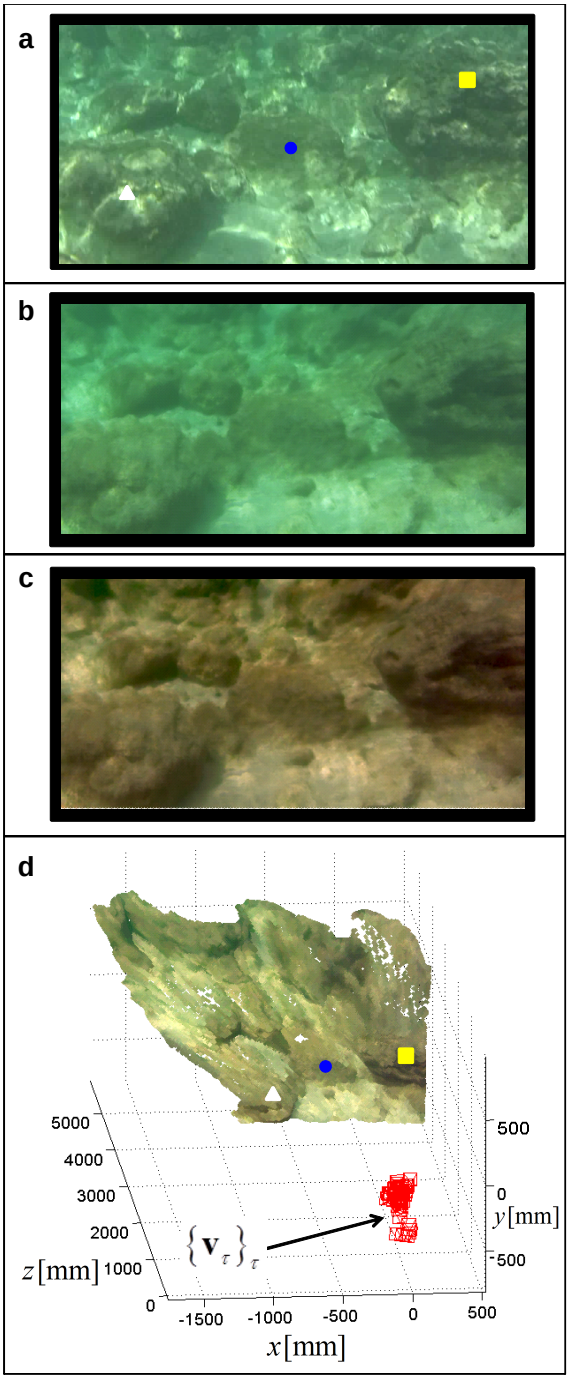

Figure 11. An underwater experiment conducted in Dor beach, Israel. [a] A frame from the left viewpoint. [b] A deflickered image rendered using temporal median on a stabilized video sequence. [c] A descattered image of [b]. [d] Using [c] for texture mapping of the estimated structure.

8. Discussion

This work presents an approach to estimate structure and motion in stereo under dynamic spatiotemporal illumination. A major principle is to use a stereoscopic range map rather than the original frames for motion estimation. Deflickered and descattered images can then be rendered.

Hence, underwater vision challenges such as low visibility and inhomogeneous lighting are partly overcome.

An assumption of rigid motion between frames is used. This assumption is invalid when capturing generally moving non-rigid objects. Therefore, it would be interesting to expand the method to detect and handle such objects as well.

Acknowledgements

We thank Marina Alterman, Yuval Bahat and Hani Swirski for their help in conducting the underwater experiments. We thank Reinhard Koch and Anne Sedlazeck for useful discussions. Yoav Schechner is a Landau Fellow - supported by the Taub Foundation. The work was supported in part by the Israel Science Foundation (Grant 1467/12) and the VPR Fund of the Technion. This work was conducted in the Ollendorff Minerva Center. Minerva is funded through the BMBF.

References

[1] M. Alterman and Y. Y. Schechner and P. Perona and J. Shamir. Detecting motion through dynamic refraction. *IEEE Trans. PAMI*, 35:245–251, 2013.

[2] M. Alterman and Y. Y. Schechner and Y. Swirski. Triangulation in random refractive distortions. In *Proc. IEEE ICCP*, 2013.

[3] S. Battiato, G. Gallo, G. Puglisi, and S. Scellato. SIFT features tracking for video stabilization. In *Proc. ICIAP*, 2007.

[4] V. Chari and P. Sturm. Multi-view geometry of the refractive plane. In *Proc. BMVC*, 2009.

[5] 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.

[6] A. Donate and E. Ribeiro. Improved reconstruction of images distorted by water waves. In *Proc. INSTICC*, 2006.

[7] R. Fattal. Single image dehazing. In *Proc. SIGGRAPH*, 2008.

[8] A. Fournier and W. T. Reeves. A simple model of ocean waves. In *Proc. SIGGRAPH*, 1986.

[9] M. N. Gamito and F. K. Musgrave. An accurate model of wave refraction over shallow water. *Computers and Graphics*, 26:291–307, 2002.

[10] J. Gedge, M. Gong, and Y. H. Yang. Refractive epipolar geometry for underwater stereo matching. In *Proc. CRV*, 2011.

[11] 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.

[12] M. Gupta, A. Agrawal, A. Veeraraghavan and S. G. Narasimhan. Structured light 3D scanning in the presence of global illumination. In *Proc. IEEE CVPR*, 2011.

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

[14] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. In *Proc. IEEE CVPR*, 2009.

[15] M. Holroyd, J. Lawrence, and T. Zickler. A coaxial optical scanner for synchronous acquisition of 3D geometry and surface reflectance. In *Proc. SIGGRAPH*, 2010.

[16] M. Irani, B. Rousso, and S. Peleg. Recovery of ego-motion using image stabilization. In *Proc. IEEE CVPR*, 1994.

[17] N. G. Jerlov. *Marine Optics*, chapter 6. Elsevier, Amsterdam, 1976.

[18] A. Jordt-Sedlazeck and R. Koch. Refractive calibration of underwater cameras. In *Proc. ECCV*, 2012.

[19] N. Joshi and M. F. Cohen. Seeing mt. rainier: Lucky imaging for multi-image denoising, sharpening, and haze removal. In *Proc. ICCP*, 2010.

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

[21] S.-J. Ko, S.-H. Lee, S.-W. Jeon, and E.-S. Kang. Fast digital image stabilizer based on gray-coded bit-plane matching. *IEEE Trans. Consum Electron*, 45(3):598 –603, 1999.

[22] 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.

[23] S. J. Koppal, S. Yamazaki, and S. G. Narasimhan. Exploiting dlp illumination dithering for reconstruction and photography of high-speed scenes. *IJCV*, 96:125–144, 2012.

[24] A. Kushnir and N. Kiryati. Shape from unstructured light. In *3DTV07*, 2007.

[25] K. J. Lee, Q. Zhao, X. Tong, M. Gong, S. Izadi, S. U. Lee, P. Tan, and S. Lin. Estimation of intrinsic image sequences from image+depth video. In *Proc. ECCV*, 2012.

[26] F. Liu, M. Gleicher, J. Wang, H. Jin, and A. Agarwala. Subspace video stabilization. *ACM TOG*, 30(1):4:1–4:10, 2011.

[27] S. Liu, Y. Wang, L. Yuan, J. Bu, P. Tan, and J. Sun. Video stabilization with a depth camera. In *Proc. IEEE CVPR*, 2012.

[28] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[29] D. K. Lynch and W. Livingston. *Color and* L*ight in* N*ature*, chapter 3,4. Cambridge U. Press, 2nd edition, 2001.

[30] W. N. McFarland and E. R. Loew. Wave produced changes in underwater light and their relations to vision. *Env. Biol. Fish.*, 8(3-4):173–184, 1983.

[31] S. G. Narasimhan, S. K. Nayar, B. Sun, and S. J. Koppal. Structured light in scattering media. In *Proc. IEEE ICCV*, 2005.

[32] E. Nascimento, M. Campos, and W. Barros. Stereo based structure recovery of underwater scenes from automatically restored images. In *Proc. SIBGRAPI*, 2009.

[33] S. Nayar and S. Narasimhan. Vision in bad weather. In *Proc. IEEE ICCV*, 1999.

[34] K. Nishino and K. Ikeuchi. Robust Simultaneous Registration of Multiple Range Images. In *Proc. ACCV*, 2002.

[35] S. Rusinkiewicz and M. Levoy. Efficient variants of the ICP algorithm. In *Proc. 3DIM*, 2001.

[36] F. Sadlo, T. Weyrich, R. Peikert, and M. Gross. A practical structured light acquisition system for point-based geometry and texture. In *Proc. IEEE Eurographics*, 2005.

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

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

[39] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *Proc. IEEE CVPR*, 2003.

[40] Y. Y. Schechner and N. Karpel. Attenuating natural flicker patterns. In *Proc. MTS/IEEE Oceans*, 2004.

[41] Y. Y. Schechner. Inversion by P^4: polarization-picture post-processing. *Phil. Trans. R. Soc. B*, 366:638–648, 2011.

[42] Y. Y. Schechner. Stereo from Flicker webpage. [Online] Available: `http://webee.technion.ac.il/~yoav/research/flicker.html`.

[43] A. Segal, D. Haehnel and S. Thrun. Generalized-ICP. In *Proc. RSS*, 2009.

[44] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: Exploring image collections in 3D. In *Proc. SIGGRAPH*, 2006.

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

[46] Y. Swirski, Y. Y. Schechner, B. Herzberg, and S. Negahdaripour. Stereo from flickering caustics. In *Proc. IEEE ICCV*, 2009.

[47] Y. Swirski, Y. Y. Schechner, and T. Nir. Variational stereo in dynamic illumination. In *Proc. IEEE ICCV*, 2011.

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

[49] T. Treibitz and Y. Y. Schechner. Instant 3Descatter. In *Proc. IEEE CVPR*, 2006.

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

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

[52] T. C. Yao-Jen Chang. Multi-view 3D reconstruction for scenes under the refractive plane with known vertical direction. In *Proc. IEEE ICCV*, 2011.

[53] G. Zeng, S. Paris, L. Quan, and F. Sillion. Surface reconstruction by propagating 3D stereo data in multiple 2D images. *IEEE Trans. PAMI*, 29(1):141–158, 2007.

[54] H. Zhang and S. Negahdaripour. Epiflow - a paradigm for tracking stereo correspondences. *CVIU*, 111(3):307–328, 2008.

[55] L. Zhang, B. Curless, and S. Seitz. Spacetime stereo: shape recovery for dynamic scenes. In *Proc. IEEE CVPR*, 2003.

[56] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *Proc. IEEE ICCV*, 1999.