

# DIRECT ESTIMATION OF THE RELATIVE ORIENTATION IN UNDERWATER ENVIRONMENT

Bashar Elnashef\*, Sagi Filin

Mapping and Geo-Information Engineering, Technion – Israel Institute of Technology, Haifa, Israel  
(Basharnashef, filin)@technion.ac.il

Commission IIWG 9

**KEY WORDS:** Underwater stereo, Pose estimation, Flat-refractive, Relative orientation, Plücker coordinates

## ABSTRACT:

While accuracy, detail, and limited time on site make photogrammetry a valuable means for underwater mapping, the establishment of reference control networks in such settings is oftentimes difficult. In that respect, the use of the coplanarity constraint becomes a valuable solution as it requires neither knowledge of object space coordinates nor setting a reference control network. Nonetheless, imaging in such domains is subjected to non-linear and depth-dependent distortions, which are caused by refractive media that alter the standard single viewpoint geometry. Accordingly, the coplanarity relation, as formulated for the in-air case does not hold in such environment and methods that have been proposed thus far for geometrical modeling of its effect require knowledge of object-space quantities. In this paper we propose a geometrically-driven approach which fulfills the coplanarity condition and thereby requires no knowledge of object space data. We also study a linear model for the establishment of this constraints. Clearly, a linear form requires neither first approximations nor iterative convergence scheme. Such an approach may prove useful not only for object space reconstruction but also as a preparatory step for application of bundle block adjustment and for outlier detection. All are key features in photogrammetric practices. Results show that no unique setup is needed for estimating the relative orientation parameters using the model and that high levels of accuracy can be achieved.

## 1. INTRODUCTION

Relative orientation enables the estimation of a minimal set of parameters that are necessary to establish coplanarity among corresponding points between two views (Mullen, 2004). Its simplified form, which requires no knowledge of object space coordinates, makes it useful as a means to obtain scene reconstruction, up to a similarity transformation, without setting a reference control network (Stewenius et al., 2006; Pollefeys and Van Gool, 1997; Hemayed, 2003; Hartley and Zisserman, 2003). These advantages become paramount in underwater environments where reference control networks are difficult to establish and where local ones become the more practical solution. Accordingly, the use of relative orientation has been applied in a diverse set of applications, including: seabed mapping, archaeological surveys, marine biology studies, vehicle navigation, or industrial equipment inspection, to name only a few (Eustice et al., 2005; Allotta et al., 2015; Ricci et al., 2015; Teixeira et al., 2016; Ludvigsen and Sørensen, 2016; Pergent et al., 2017; Herkül et al., 2017).

Establishing a relative orientation in underwater environments is challenged, however, by limited visibility due to scattering and absorbance of light, and by refraction that alters light rays trajectory from their standard 'collinear' path and ultimately the coplanarity relation (Menna et al., 2017). The problem of recovering 3-D geometry in the presence of refraction has been studied extensively in recent years. The majority of approaches rely on standard in-air models, assuming that the pinhole camera model with distortion can compensate for refraction (e.g., Pizarro et al., 2003; Singh et al., 2007; Shortis, 2015). These models have proved applicable, but analysis of the actual image-point correction due to refraction shows that for

close-range imaging, the actual aberration is depth dependent, and three dimensional (Sedlazeck and Koch, 2012; Shortis, 2015). Modeling the effect of imaging through refractive media has shown that the standard single viewpoint perspective (SVP) model no longer applies in this setting and introduces systematic errors that are proportional to the distance between the entrance pupil and the flat port (Menna et al., 2017). To handle these effects, recent studies considered the introduction of an explicit modeling of the refraction effect. As an example, Chaudhury et al. (2015) implemented a ray-tracing based model that allows expressing the underlying refractive geometry as an extension of projective geometry. The authors assumed, however, that refraction occurs only at the camera center, allowing to geometrically estimate the relationship between the observed image point and the one obtained as if refraction did not occur. The case of a fixed interface and the moving camera was considered by Chen et al. (2011) where the authors assumed that the imaging system is integrated with an inertial measurement unit (IMU) so that the pitch and yaw angles of the camera are considered known. The authors then proposed a closed-form solution to the absolute orientation estimation, yet required the vertical displacement of the camera to be known in advance. A setup of two cameras embedded in different watertight housing was considered in Kang et al. (2012). The authors neglected the glass thickness and considered the image plane and the glass interface to be parallel. They estimated then the air medium thickness for each imaging system and the five relative pose parameters in a non-linear optimization procedure. Telem and Filin (2013) derived a coplanarity constraint, which is based on the varifocal representation. The authors demonstrated that because of refraction, the linear relation between views does not hold when imaging underwater, leading to a non-linear and inhomogeneous coplanarity constraint. Aiming to accommodate for a thick glass interface found in deep-sea high pressure

\*Corresponding author

housings, Jordt et al. (2016) proposed a ray-tracing refractive structure-from-motion (SfM) framework by extending the one previously proposed in Jordt-Sedlazeck et al. (2013) and Jordt-Sedlazeck and Koch (2013). The authors relied on the coplanarity constraint to estimate the relative camera motion between two successive views while using a non-linear refractive bundle adjustment. The proposed refractive SfM required a good initialization, especially for the two-view pose estimation case, and relied on a computationally demanding genetic optimization. Furthermore, the model exhibited sensitivity to noise, especially when using thin glass interface or when the distance between the perspective center and the interface was small. A pose and calibration models for an underwater stereo imaging system was proposed by Zhang et al. (2018) where the authors remodeled the underwater imaging system in the view of the light filled parametrization and proposed a forward projection error function which was minimized in a non-linear optimization fashion.

The literature shows that because of refraction, underwater imaging suffers from depth-dependent, non-linear distortions and that the imaging system does not follow the standard SVP model. To handle that, recent research followed ray-tracing based principles and incorporated robust optimization strategies, proposed to algorithmically reduce the measurement noise, or used hardware-driven solutions. While exerting high computational demand, they still fall short of obtaining satisfactory levels of accuracies and are often supplemented by elaborate setups, or require an a priori knowledge of part of the camera pose parameters. In this paper, we propose a geometrically driven model that accounts for the refraction effect and constitutes an alternative formulation of the relative orientation and the coplanarity constraint. We show that the pose parameters can be estimated linearly and directly, with no need for first approximations or iterations. The linearity is achieved by separating the relative pose from the underwater-related system parameters where the refraction effect is manifested. The advantages of our model are the following: firstly, the relative pose parameters can be estimated linearly; secondly, and contrary to the in-air case, the model allows to estimate the scale of the system; and thirdly, 3-D data can be extracted with no need to establish an explicit reference frame. The model is analyzed by simulated experiments and in actual underwater conditions. Evaluation is of the inner and geometric accuracy, both are analyzed by the reconstructed point-sets. Results show that a high-level of accuracy is reached while facilitating a flexible orientation strategy for modeling in underwater environments.

## 2. METHODOLOGY

### 2.1 Varifocal model

We consider an imaging system made of a camera shielded by a flat-port housing. The image formation model is of a ray that propagates from an in-water point until it refracts at the port glass-interface and then propagates through the air until it reaches the image plane while passing through the perspective center (Fig. 1). The system is defined by the camera's focal length,  $f$ , the distance from the housing interface to the perspective center,  $F$ , and the refractive indices. Snell's law of refraction suggests that the refracted ray, the normal to the interface surface, and the incoming- and outgoing-rays, all lie on the same plane (Fig. 1). As the refracted rays pass through the perspective center, all planes of refraction revolve about the optical axis as long as it coincides with the normal to the

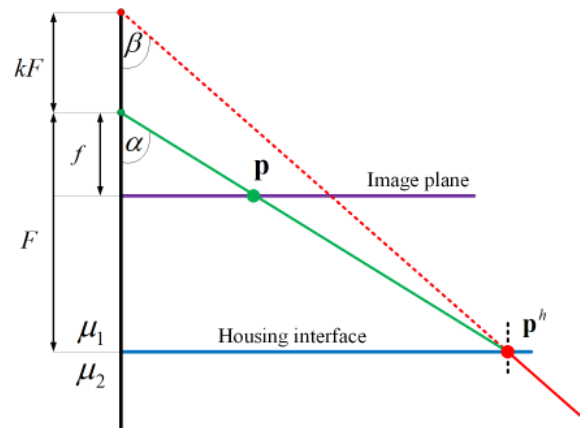


Figure 1: Varifocal configuration, the dashed line follows the direction of the underwater ray until intersecting the system axis, where  $kF$  is the perspective center offset,  $\alpha$  is the angle of refraction; and  $\beta$  is the angle of incidence ( $\mu_1$  and  $\mu_2$  the refractive indices of air and water, respectively).

interface. Hence, the axiality of the system. Axiality is also a property of the varifocal model, but here the position of the perspective center is modified along the optical axis to maintain the collinearity of the incoming ray.

**Axial camera modeling** – Modifying the imaging system with respect to the interface (Fig. 1),  $F$  becomes the principal distance, and coordinates are given by:

$$\bar{\mathbf{x}} = [\bar{x}_i, \bar{y}_i, -(F + k_i F)]^T \quad (1)$$

where  $\bar{x}_i = \frac{F}{f} x_i$  and  $\bar{y}_i = \frac{F}{f} y_i$  are the housing system coordinates for a given image plane coordinates,  $x_i$  and  $y_i$ .

The collinearity form in this system can be written as:

$$\begin{bmatrix} \bar{x}_i \\ \bar{y}_i \\ -F \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ -k_i F \end{bmatrix} = \lambda_i \begin{bmatrix} \mathbf{r}_1 \mathbf{X}_i - t_x \\ \mathbf{r}_2 \mathbf{X}_i - t_y \\ \mathbf{r}_3 \mathbf{X}_i - t_z - k_i F \end{bmatrix} \quad (2)$$

where  $\mathbf{X}_i$  is the 3-D coordinate of a point in object space;  $\mathbf{r}_i$  is the  $i$ -th row of the rotation matrix  $\mathbf{R}$ ;  $t_x$ ,  $t_y$ , and  $t_z$  are the camera position in the camera reference frame;  $\lambda_i$  is a scale factor; and  $k_i$ :

$$k_i(\alpha_i) = \sqrt{\frac{1 - \mu^2 \sin^2 \alpha_i}{\mu^2 \cos^2 \alpha_i}} - 1 \quad (3)$$

is the principal distance adaptation to the refraction effect, where  $\alpha$  is the angle of incidence (Fig. 1); and  $\mu = \mu_1/\mu_2$  is the refractive index ratio between the indices of air and water.

**Deviation from parallelism** – The derivation of Eq. (2) assumed that the optical axis and the normal to the interface coincide. This assumption fails to hold when the image plane is not parallel to the housing interface. In such a case all the planes of refraction still pass through the perspective center. They do not coincide however with the optical axis, rather with the normal to the interface, which is common to all planes of refraction. The offset between the normal to the interface and the optical axis is an image related term which is defined by two rotation angles, one off the optical axis and another about it. Telem and Filin (2010) have shown that the rotation about the  $z$ -axis is strongly correlated with the exterior orientation angles, thus ab-

sorbed, and cannot be estimated independently. Consequently, the transformation between the two systems can be approximated by the use of a single rotation about the y-axis. Hence, the modification becomes:

$$\mathbf{x}' = \mathbf{R}_y(\eta) \mathbf{x} \quad (4)$$

where  $\mathbf{R}_y(\eta)$  is the rotation matrix from the image plane to the interface by an angle  $\eta$ . Thus, the modified expression for  $k$  becomes:

$$\tilde{k}(\mathbf{x}) = \frac{1}{\mu} \sqrt{\frac{\mathbf{x}^T \mathbf{x}}{\mathbf{x}^T \mathbf{M}_3 \mathbf{x}} + \mu^2 \left( \frac{\mathbf{x}^T (\mathbf{M}_3 - \mathbf{I}) \mathbf{x}}{\mathbf{x}^T \mathbf{M}_3 \mathbf{x}} \right)} - 1 \quad (5)$$

where  $\mathbf{x}$  is the image point coordinate, and  $\mathbf{M}_3 = \mathbf{r}_3 \otimes \mathbf{r}_3$  is the outer product of the third row in  $\mathbf{R}_y(\eta)$  in which the refraction effect is accounted for. As offsetting the perspective center should be performed in reference to the system in which the normal to the interface acts as the optical axis, both image and object-space related vectors should transform to this system, followed by a modification to the perspective center:

$$\mathbf{M} \begin{pmatrix} \bar{x}_i \\ \bar{y}_i \\ -F \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ \tilde{k}_i F \end{pmatrix} = \lambda_i \left[ \mathbf{M} \begin{pmatrix} \mathbf{r}_1 \mathbf{X}_i - t_x \\ \mathbf{r}_2 \mathbf{X}_i - t_y \\ \mathbf{r}_3 \mathbf{X}_i - t_z \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ \tilde{k}_i F \end{pmatrix} \right] \quad (6)$$

The glass interface effect on the system was not considered to that point, but it was demonstrated by [Telem and Filin \(2010\)](#) that for a standard underwater imaging system it introduces a lateral shift which is smaller by an order of magnitude or more than the one caused due to angular refraction. This shift can be reduced to a constant value, factored by the glass thickness. Hence, it is absorbed by  $F$ .

## 2.2 Plücker coordinates of a 3-D line

To establish the relative orientation we define the ray's direction using Plücker representation of lines. We show that such a representation facilitates a linear estimation of the camera pose parameters. We represent the projected image rays as 3-D lines using Plücker coordinates. Of the definitions for a Plücker 3-D line we use the vector form ([Förstner and Wrobel, 2016](#)). Let  $\mathbf{A}$  and  $\mathbf{B}$  be homogeneous coordinates of two 3-D points defining a line. The Plücker line representation is given by:

$$\mathbf{L} = \begin{pmatrix} A_4 B_1 - A_1 B_4 \\ A_4 B_2 - A_2 B_4 \\ A_4 B_3 - A_3 B_4 \\ A_3 B_2 - A_2 B_3 \\ A_1 B_3 - A_3 B_1 \\ A_2 B_1 - A_1 B_2 \end{pmatrix} \quad (7)$$

where  $A_i, B_i$  is the  $i^{\text{th}}$  ordinate of the respective point.  $\mathbf{L}$  can be split into two 3-vectors  $\mathbf{a}$  and  $\mathbf{b}$ ,

$$\mathbf{L} = (\mathbf{a}^T \quad \mathbf{b}^T)^T = (L_1 \quad L_2 \quad L_3 \quad L_4 \quad L_5 \quad L_6)^T \quad (8)$$

where,  $\mathbf{a}$  and  $\mathbf{b}$  are the direction and moment of the 3-space line, respectively. To define a proper line in space these two sub vectors must satisfy the constraint:

$$\mathbf{a}^T \mathbf{b} = 0 \quad (9)$$

Considering two 3-space lines,  $\mathbf{L}$  and  $\mathbf{L}'$ , a coplanarity relation between them is obtained if and only if they fulfill the following constraint:

$$\mathbf{L}^T \begin{pmatrix} \mathbf{0}_{3 \times 3} & \mathbf{I}_{3 \times 3} \\ \mathbf{I}_{3 \times 3} & \mathbf{0}_{3 \times 3} \end{pmatrix} \mathbf{L}' = \mathbf{L}^T \mathbf{W} \mathbf{L}' = 0 \quad (10)$$

where,  $\mathbf{W}$  is a  $6 \times 6$  permutation matrix.

**Transformations** – a metric transformation defined by a rotation matrix  $\mathbf{R}$  and a translation vector  $\mathbf{t}$ , acts on a point  $\mathbf{X}$  by:

$$\mathbf{X} \rightarrow \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \mathbf{X} \quad (11)$$

Accordingly, Plücker line coordinates are transformed by,

$$\mathbf{L} \rightarrow \begin{pmatrix} \mathbf{R} & \mathbf{0} \\ -[\mathbf{t}]_{\times} \mathbf{R} & 1 \end{pmatrix} \mathbf{L} \quad (12)$$

where,  $[\mathbf{t}]_{\times}$  is a skew symmetric matrix ([Förstner and Wrobel, 2016](#)).

## 2.3 Underwater relative orientation

Considering,  $\mathbf{L}$  and  $\mathbf{L}'$  as lines that represent the rays projected from two cameras meeting in a point in object-space, Eq. (10) can be written as:

$$\mathbf{L}^T \begin{pmatrix} [\mathbf{t}]_{\times} \mathbf{R} & \mathbf{R} \\ \mathbf{R} & \mathbf{0} \end{pmatrix} \mathbf{L}' = 0 \quad (13)$$

Describing a ray by a point  $\mathbf{v}$  that lies on it and a unit vector direction  $\hat{\mathbf{x}}$ . The Plücker coordinates of two lines corresponding to image pair correspondence  $i$  are given by,

$$\mathbf{L}_i = \begin{pmatrix} \hat{\mathbf{x}}_i \\ \mathbf{v}_i \times \hat{\mathbf{x}}_i \end{pmatrix} \quad ; \quad \mathbf{L}'_i = \begin{pmatrix} \hat{\mathbf{x}}'_i \\ \mathbf{v}'_i \times \hat{\mathbf{x}}'_i \end{pmatrix} \quad (14)$$

In our case,  $\hat{\mathbf{x}}_i$  would be the direction defined by the corrected perspective center  $\mathbf{v}_i = [0, 0, -k_i F]$  and its corresponding housing system point  $\bar{\mathbf{x}}_i$  (Eq. 1) in the first image. Setting the point on the line to be  $\mathbf{v}_i$ , Eq. (14) can be written in terms of the varifocal model as follows,

$$\mathbf{L}_i = (\bar{x}_i \quad \bar{y}_i \quad F + F\tilde{k}_i \quad -\tilde{k}_i \bar{y}_i \quad \tilde{k}_i \bar{x}_i \quad 0)^T \quad (15)$$

As the last element of Plücker rays is zero, the  $6 \times 6$  transformation matrix in (Eq. 13) can be reduced to the following  $5 \times 5$  matrix:

$$\mathbf{E}_r = \begin{pmatrix} e_{11} & e_{12} & e_{13} & r_{11} & r_{12} \\ e_{21} & e_{22} & e_{23} & r_{21} & r_{22} \\ e_{31} & e_{32} & e_{33} & r_{31} & r_{32} \\ r_{11} & r_{12} & r_{13} & 0 & 0 \\ r_{21} & r_{22} & r_{23} & 0 & 0 \end{pmatrix} \quad (16)$$

yielding the constraint:

$$\tilde{\mathbf{L}}_i^T \mathbf{E}_r \tilde{\mathbf{L}}'_i = 0 \quad (17)$$

where,  $\tilde{\mathbf{L}}_i$  and  $\tilde{\mathbf{L}}'_i$  are a 5-length vectors. We can also write Eq. (17) as:

$$\tilde{\mathbf{x}}_i^T \mathbf{E} \tilde{\mathbf{x}}'_i + \tilde{\mathbf{x}}_i^T \mathbf{R} (\mathbf{v}'_i \times \tilde{\mathbf{x}}'_i) + (\mathbf{v}_i \times \tilde{\mathbf{x}}_i) \mathbf{R} \tilde{\mathbf{x}}'_i = 0 \quad (18)$$

Eq. (17) provides a form from which the relative pose parameters can be estimated linearly. For that purpose, we use the sin-

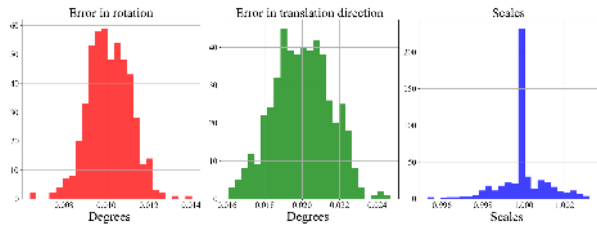


Figure 2: Histograms of the estimation accuracy using 30 corresponding points and based on 500 randomly simulated tests. In all tests, a Gaussian noise with  $\sigma = \pm 0.5$  pixels was introduced to the image measurements.

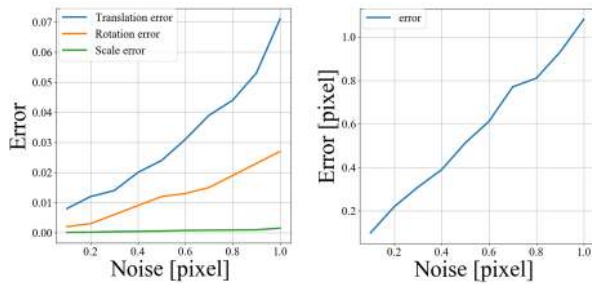


Figure 3: Pose estimation errors as a function of Gaussian noise. Left shows the estimation accuracy in rotation, translation, and scale (units are in accordance with Eqs. (19a), (19b) and (19c)) as a function of Gaussian noise level. Right shows the reprojected error.

gular value decomposition (SVD; Golub and Van Loan (1983)) based solution for  $\mathbf{E}_r$ , which yields satisfactory results. However, we propose a two-step approach, an alternative solution scheme to obtain better estimates. First the rotation matrix is estimated from  $\mathbf{E} = [\mathbf{t}]_x \mathbf{R}$ , which is computed by the first element in Eq. (18). Secondly, a non-linear refinement of this estimates is directly computed from the complete form of Eq. (18). To compute the scale, we fix the refined rotation matrix in Eq. (18) and compute the translation vector,  $\mathbf{t}$ . As  $\mathbf{t}$  is the only unknown and Eq. (18) is a non-homogeneous set of equation (unless  $\mathbf{R} = \mathbf{I}$ , meaning no motion), thus  $\mathbf{t}$  can be computed using linear least-squares adjustment, with no scale ambiguity.

### 3. EXPERIMENTS

Evaluation of the model was performed by studying two scenarios, one in which the pose parameters were estimated and compared to ground truth, and another in which reconstruction error was evaluated. The evaluation was carried out using simulations with settings that resemble actual underwater imaging conditions and using real-world experiments. For both real and simulated experiments, the proposed model performance was characterized by the following metrics: (i) angle difference between the rotation axes of  $\mathbf{R}$  and  $\hat{\mathbf{R}}$  (Eq. 19a); (ii) the direction difference between  $\mathbf{t}$  and  $\hat{\mathbf{t}}$  (Eq. 19b); and (iii) the scale error between  $\mathbf{t}$  and  $\hat{\mathbf{t}}$  (Eq. 19c).

$$\delta_\theta = \cos^{-1} \left( \frac{\text{tr}(\mathbf{R}\hat{\mathbf{R}}^T) - 1}{2} \right) \quad (19a)$$

$$\delta_{\mathbf{t}} = \cos^{-1} \left( \frac{\mathbf{t}^T \hat{\mathbf{t}}}{\|\mathbf{t}\| \|\hat{\mathbf{t}}\|} \right) \quad (19b)$$

$$\epsilon_{scale} = \left| 1 - \frac{\|\mathbf{t}\|}{\|\hat{\mathbf{t}}\|} \right| \quad (19c)$$

The parameters  $\mathbf{R}$ ,  $\mathbf{t}$  are the ground-truth information used for the simulations, and  $\hat{\mathbf{R}}$ ,  $\hat{\mathbf{t}}$  correspond to the estimated equivalents. Note that all these matrices or vectors are defined in absolute scales rather than up to a scale as is in the in-air case.

#### 3.1 Synthetic simulations

To test the pose parameter estimation model we consider an imaging system with a standard frame camera and a flat-port housing. The settings consisted of a  $3000 \times 2000$  pixels frame camera with a  $7.5 \mu\text{m}$  pixel size; a 24 mm lens;  $\mu_{glass} = 1.5$ ;  $\mu_{water} = 1.333$ ;  $F = 50$  mm; and  $\eta = 1^\circ$ . A normally distributed additive random noise with standard deviation ranging between  $\sigma = [0.1, 1]$  pixels was introduced to the image coordinates.

The first experiment tested the estimation accuracy for rotation, translation, and scale in the presence of  $\sigma = \pm 0.5$  pixels of Gaussian noise. Histograms of the estimation accuracy of different metrics (Eqs. 19a, 19b and 19c) are plotted in Fig. (2). The second experiment tested the quality and stability of the camera pose parameters (position and orientation), and scale estimation in the presence of Gaussian noise. Here measurement noise varied from 0.1 to 1 pixels. The accuracy estimates are discussed only for  $\sigma = \pm 0.5$ , other levels are plotted in the designated graphs in Fig. (3). The application of the proposed model for that noise level yielded an angular error of  $\pm 0.012^\circ$ . The scale estimation accuracy was  $5.5e^{-4}$ . The reprojected error shows a linear trend with the increase of the noise level (Fig. 3), an indication of the stability of the model. In all experiments convergence was reached within three iterations, one for the estimation of the approximated values and two others for the estimation by the optimal form in Eq. (18).

#### 3.2 Real world – Open sea experiment

In addition to the synthetic evaluation, another test was performed under real-world conditions in open sea. This environment provided less controlled and different settings. Here a Nikon D70s with a fixed 24 mm lens shielded by an IKELITE housing, and two 2-D rigid plates were immersed in water, both are used for testing the quality of the pose estimation. A dry calibration phase was performed while the camera was in the housing (Elnashef and Filin, 2019), then, a wet calibration was carried out using a total of five images. The a posteriori standard deviation of the adjustment was  $\hat{\sigma}_0 = \pm 0.42$  pixels.  $F$  was estimated with an accuracy of  $\pm 0.07$  mm,  $\pm 1e^{-6}$  for  $\mu$ , and  $\pm 0.05^\circ$  for  $\eta$ . The system parameters were then fixed in the next stage. We used ORB key-point detector for feature detection, then a matching between correspondence was performed to find corresponding points between the two images shown in (Fig. 4). Note that, the two images used in the relative orientation estimation were different than those used in the calibration stage.

Before approaching the non-linear estimation, approximate pose parameters were computed. To obtain them, we estimated the essential matrix,  $\mathbf{E} = [\mathbf{t}]_x \mathbf{R}$  using the housing system coordinates. Hence, we solved for  $\mathbf{E} = \bar{\mathbf{x}}_i^T \mathbf{E} \bar{\mathbf{x}}'_j$ , which offers adequate approximation for Eq. (18). We applied the standard 5-point and the RanSac algorithms (Stewenius et al., 2006) to extract the parameters. With these values, we then optimally minimized the geometric error, by directly solving for the three rotations angles ( $\omega, \phi, \kappa$ ) and two translation ( $t_x, t_y$ ) (Eq. 18). Finally, to estimate the translation vector with scale, we first computed the optimal rotation matrix  $\hat{\mathbf{R}}$  from the estimated angles (Eq. 18). Then, fixing it in Eq. (18), we solved only for the

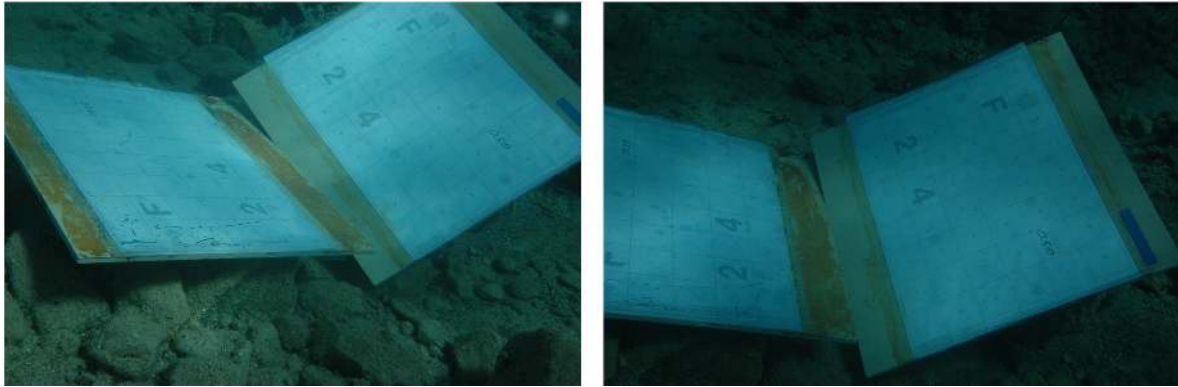


Figure 4: Images taken of the 2D plates in an open-sea preeminent.

three translation parameters  $\hat{\mathbf{t}} = [t_x, t_y, t_z]$ , using a standard least-squares solution.

**Testing 3-D point measurement accuracy** – We evaluated the application of these parameters when applied to the measurement of 3-D object-space points using the varifocal model proposed by Elnashef and Filin (2019). Measurements were performed on tilted test-plate (with respect to the camera) by two images. The evaluated measures included coplanarity, collinearity, parallelism and orthogonality of the measured grid lines, and measurement of the grid dimensions. All 63 corners of the 7 cm-spaced, 42×56 cm, grid (Fig. 4) were measured. The first test evaluated the coplanarity of the mapped points, and a plane that was fitted to them yielded a  $\pm 0.15$  mm deviation. The collinearity test for points lying on straight-lines, and as a consequence the removal of the underwater effect that bends them, yielded a mean deviation of  $\pm 0.21$  mm and a maximal deviation of 0.36 mm from collinearity. Computation of the angles between parallel and orthogonal lines shows an angular error of 47" and 27" in parallelism and orthogonality, respectively. These results are equivalent to 0.54 mm and 0.42 mm deviation, with respect to their actual length on the target (56 cm and 42 cm, respectively). Such deviations are comparable in their accuracies to the ones reached in our other experiments.

#### 4. CONCLUSIONS

This paper proposed a geometrical driven and scaled-relative-orientation model for the underwater environment. Its main advantage lies in the preservation of the coplanarity constraint with no need to gain explicit knowledge of object point coordinates, while also separating the pose parameters from the refraction related ones. Therefore, our model enables the estimation of the underwater related relative orientation parameters with the addition of the scale. The experiments demonstrated how estimation of the orientation parameters yielded high levels of accuracy, robustness to first approximations, and high level of noise. The linear estimation form requires 17 point-correspondences, which may limit its applicability in images with low number of correspondence, and would increase the computational demand in the presence of high fraction of outliers. This was alleviated, however, by the direct estimation methodology proposed in the paper, which required only 5 points and allowed us to solve the pose parameters in an optimal manner by minimization of the reprojection error. We showed that by applying our proposed relative orientation model, where scale is also estimable, it was possible to obtain high estimation and also reconstruction accuracy in close-range underwater

imaging with no domain knowledge given.

#### 5. ACKNOWLEDGMENT

This work was supported in part by the Israel Ministry of Science, Technology and Space grant # 3-12487.

#### REFERENCES

- Allotta, B., Costanzi, R., Ridolfi, A., Colombo, C., Bellavia, F., Fanfani, M., Pazzaglia, F., Salvetti, O., Moroni, D., Pascali, M. A. et al., 2015. The arrows project: adapting and developing robotics technologies for underwater archaeology. *International Federation of Automatic Control (IFAC)*, 194–199.
- Chaudhury, S., Agarwal, T., Maheshwari, P., 2015. Multiple view 3-d reconstruction in water. *2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, IEEE, 1–4.
- Chen, Z., Wong, K., Matsushita, Y., Zhu, X., Liu, M., 2011. Self-calibrating depth from refraction. *IEEE International Conference on Computer Vision Proceedings*, IEEE. The Journal's web site is located at <http://ieeexplore.ieee.org/xpl> . . . .
- Elnashef, B., Filin, S., 2019. Direct linear and refraction-invariant pose estimation and calibration model for underwater imaging. *ISPRS Journal of Photogrammetry and Remote Sensing*, 154, 259–271.
- Eustice, R., Camilli, R., Singh, H., 2005. Towards bathymetry-optimized doppler re-navigation for auvs. *OCEANS, 2005. Proceedings of MTS/IEEE*, 1430–1436.
- Förstner, W., Wrobel, B. P., 2016. *Photogrammetric computer vision*. Springer.
- Golub, G., Van Loan, C., 1983. *Matrix Computations*. Johns Hopkins Studies in Atlantic History & Culture, Johns Hopkins University Press.
- Hartley, R., Zisserman, A., 2003. *Multiple view geometry in computer vision*. Cambridge university press.
- Hemayed, E. E., 2003. A survey of camera self-calibration. *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, 2003.*, IEEE, 351–357.

- Herkül, K., Peterson, A., Paekivi, S., 2017. Applying multibeam sonar and mathematical modeling for mapping seabed substrate and biota of offshore shallows. *Estuarine, Coastal and Shelf Science*, 192, pp. 57–71.
- Jordt, A., Köser, K., Koch, R., 2016. Refractive 3D reconstruction on underwater images. *Methods in Oceanography*, 15, pp. 90–113.
- Jordt-Sedlazeck, A., Jung, D., Koch, R., 2013. Refractive plane sweep for underwater images. *German Conference on Pattern Recognition*, Springer, pp. 333–342.
- Jordt-Sedlazeck, A., Koch, R., 2013. Refractive structure-from-motion on underwater images. *Proceedings of the IEEE international Conference on Computer Vision*, 57–64.
- Kang, L., Wu, L., Yang, Y.-H., 2012. Experimental study of the influence of refraction on underwater three-dimensional reconstruction using the svp camera model. *Applied optics*, 51(31), 7591–7603.
- Ludvigsen, M., Sørensen, A. J., 2016. Towards integrated autonomous underwater operations for ocean mapping and monitoring. *Annual Reviews in Control*, 42, pp. 145–157.
- Menna, F., Nocerino, E., Remondino, F., 2017. Flat versus hemispherical dome ports in underwater photogrammetry. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42(2W3), pp. 481–487.
- Mullen, R., 2004. *Manual of photogrammetry*. Asprs Publications.
- Pergent, G., Monnier, B., Clabaut, P., Gascon, G., Pergent-Martini, C., Valette-Sansevin, A., 2017. Innovative method for optimizing Side-Scan Sonar mapping: The blind band unveiled. *Estuarine, Coastal and Shelf Science*, 194, pp. 77–83.
- Pizarro, O., Eustice, R. M., Singh, H., 2003. Relative pose estimation for instrumented, calibrated imaging platforms. *DICTA*, 601–612.
- Pollefeys, M., Van Gool, L., 1997. Self-calibration from the absolute conic on the plane at infinity. *International Conference on Computer Analysis of Images and Patterns*, Springer, 175–182.
- Ricci, S., Antonelli, F., Perasso, C. S., 2015. Aspects of biodeterioration of lapideous submerged artefacts: 3D methodologies application. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(5), pp. 191–194.
- Sedlazeck, A., Koch, R., 2012. Perspective and non-perspective camera models in underwater imaging—overview and error analysis. *Outdoor and Large-Scale Real-World Scene Analysis*, 212–242.
- Shortis, M., 2015. Calibration techniques for accurate measurements by underwater camera systems. *Sensors*, 15(12), 30810–30826.
- Singh, H., Roman, C., Pizarro, O., Eustice, R., 2007. Advances in high resolution imaging from underwater vehicles. *Robotics Research*, 430–448.
- Stewenius, H., Engels, C., Nistér, D., 2006. Recent developments on direct relative orientation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(4), 284–294.
- Teixeira, P. V., Kaess, M., Hover, F. S., Leonard, J. J., 2016. Underwater inspection using sonar-based volumetric submaps. *Intelligent Robots and Systems (IROS), IEEE/RSJ International Conference*, 4288–4295.
- Telem, G., Filin, S., 2010. Photogrammetric modeling of underwater environments. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(5), pp. 433–444.
- Telem, G., Filin, S., 2013. Photogrammetric modeling of the relative orientation in underwater environments. *ISPRS journal of photogrammetry and remote sensing*, 86, 150–156.
- Zhang, C., Zhang, X., Zhu, Y., Li, J., Tu, D., 2018. Model and calibration of underwater stereo vision based on the light field. *Measurement Science and Technology*, 29(10), pp. 105402.