2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

Online Refractive Camera Model Calibration in Visual Inertial Odometry

Mohit Singh, and Kostas Alexis

***Abstract*— This paper presents a general refractive camera model and online co-estimation of odometry and the refractive index of an unknown media. This enables operation in diverse and varying refractive fluids, given only the camera calibration in air. The refractive index is estimated online as a state variable of a monocular visual-inertial odometry framework in an iterative formulation using the proposed camera model. The method was verified on data collected using an underwater robot traversing inside a pool. The evaluations demonstrate convergence to the ideal refractive index for water despite significant perturbations in the initialization. Simultaneously, the approach enables on-par visual-inertial odometry performance in refractive media without prior knowledge of the refractive index or requirement of medium-specific camera calibration.**

I. Introduction

Underwater robots are used in various fields including environmental surveillance [1], sea floor mapping [2], and structural inspection [3]. To operate autonomously, these systems typically rely on a variety of domain-specific sensors [4], including 3D sonars, acoustics, and Doppler Velocity Log instruments combined with Inertial Measurement Units (IMUs), while vision cameras are used to observe areas of interest but playing a secondary role in localization [5–8]. This is due to the fact that visual data in open waters without nearby structures are of limited utility for odometry estimation. However, the growing need for close-up underwater inspection in cluttered settings, such as oil & gas facilities or shipwrecks, has led to a rising interest in vision-based underwater robots [9–14]. The affordability of cameras further contributes to this trend. Driven by these observations, research has focused on Visual-Inertial Odometry (VIO) techniques underwater [10, 11, 15, 16]. Among the efforts in the domain, it is customary to calibrate the camera model set-up underwater in the area of interest, a choice driven by phenomena such as light refraction in the water and the dependency of refractive index on factors such as temperature, salinity , pressure, wavelength, and more [17].

Motivated by the above and with the goal of eliminating the need for the laborious task of underwater calibration for any camera system with a flat-port, this work contributes *a)* online estimation of a medium's refractive index n as a state variable in a visual-inertial odometry framework for a general camera-and-IMU system thus enabling versatile underwater VIO, *b)* online rectification of observed landmarks using the estimated refractive index and camera parameters obtained from conventional calibration in air, *c)* formulation of a sensitivity heuristic for robustness against degenerate motions and noisy image conditions, as well as *d)* verified convergence of refractive index given a large range of perturbation, highlighting the ability to adapt to a wide range of fluids accessible for robotics and computer vision applications.

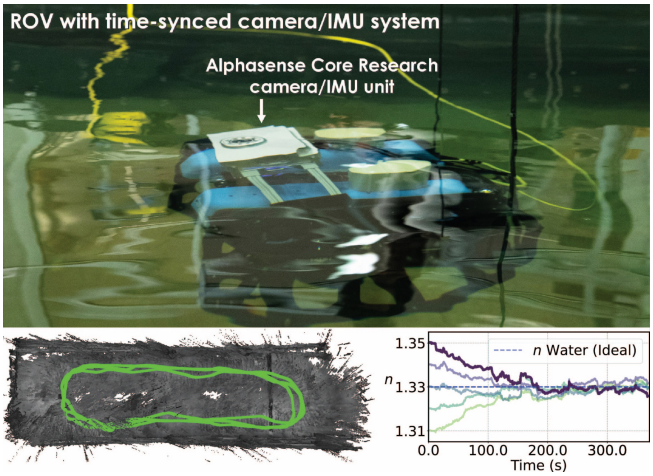

Fig. 1. Instance of the conducted experimental studies employing a Remotely Operated Vehicle integrating a time-synchronized camera/IMU system navigating in a pool subject to diverse light conditions. The proposed approach enables online estimation of the medium's refractive index and thus adjusts the camera model employed within visual-inertial odometry.

Furthermore, compared to our relevant prior work [18], this contribution *i)* presents a formulation for a monocular camera and IMU which highlights its generalizability compared to the stereo-only formulation in [18], *ii)* enables the co-estimation of the refractive index embedded into VIO instead of a 2-step approach where VIO used a rectified image from a separate refractive module, and *iii)* formulates a sensitivity heuristic as a function of the essential matrix and bearing vector instead of a stereo-specific formulation.

The presented approach was extensively verified in experimental studies to assess its performance both in terms of real-time refractive index estimation, as well as overall VIO accuracy and robustness against ground truth. To support this research, an extensive dataset featuring a Remotely Operated Vehicle (ROV) equipped with a time-synchronized 5-camera/IMU setup in a laboratory pool (Figure 1) involving diverse light conditions was collected and made publicly available. The collected data are openly released augmenting our previous release in `https://github.com/ntnu-arl/underwater-datasets`.

In the remainder of this paper, Section II presents related

This material was supported by the Research Council of Norway Award NO-327292.

The authors are with the Norwegian University of Science and Technology (NTNU), O. S. Bragstads Plass 2D, 7034, Trondheim, Norway `mohit.singh@ntnu.no`

work, while the proposed camera model is detailed in Section III and the visual-inertial co-estimation is presented in Section IV. Experimental studies are shown in Section V, while conclusions are drawn in Section VI.

II. Related Work

This contribution relates to the body of work on adaptive camera modeling, refractive index estimation and visual odometry across water and other media, alongside the underlying concepts in camera modeling and motion estimation (e.g., [19–22]). A formulation of the fundamental matrix augmented to account for refractive effects has been detailed in [23]. The study in [24] derives a plate refractive camera model as a pixel-wise variable viewpoint pinhole model, a caustic surface, a calibration methodology, and a refraction-based triangulation. The work in [25] concentrates on cameras deviating from the single viewpoint (non-SVP) architecture proposing a physics-grounded model. The authors in [26] address both perspective and non-perspective camera models and discuss the limitations of the traditional pinhole model underwater. The work in [27] introduces modeling for multiple layer flat refractive geometry and enables refractive index calculation of media with known scene geometry. The contribution in [28] presents a method for 3D reconstruction from underwater images and a bundle-adjustment technique for autocalibration contingent upon a precise initial estimate. The authors in [29] approach handling refraction through a pixel-wise varifocal model and use linear extrinsic camera calibration based on a calibrating target.

Our work also relates to efforts that advanced the application of multiple view geometry in underwater environments. In [30] the authors introduce a camera model to enable underwater scene mapping and compute the medium's refractive index while removing geometric refraction effects based on image point correspondences. The work in [31] focuses on deriving the absolute pose of a camera viewing through a known refractive plane, while emphasizing the intricacies introduced by Snell's law ambiguities. The studies in [32, 33] further advance pose estimation in refractive media.

Recent work has focused on accounting for refractive effects in visual-inertial fusion. The study in [11] focuses on improving underwater VIO by introducing an image rectification technique correcting distortions caused by both water-air refraction and camera lens issues. It employs an approximate SVP model, while calibration is performed underwater. Towards self-calibration, the work in [16] presents a method that starts from calibration in air and then estimates environmental indexes in the water under small angle approximations and thus reduced field-of-view cameras. Focusing on stereo, [34] explicitly considers a refractive camera model but relies on fiducials to enable localization. Further on stereo, [35] presents underwater localization and mapping with a refractive camera model calculated offline using calibration images and delivering nonlinear epipolar curves for stereo matching. Considering the application-specific alternative of localizing a robot inside the water by structures above the water, [36] utilizes an upward-facing stereo camera to build a global ceiling map. Outside of this niche set of works that explicitly consider refractive camera models, most approaches in underwater VIO tend to calibrate their cameras underwater with conventional methods, in shallow waters before being applied in different environments or directly within the target area [10, 11, 14, 15, 37]. Approaches may also be multi-modal (e.g., with the fusion of sonar [38] or pressure readings [39]).

III. Refractive Camera Model

Building upon our previous work in [18] this model is derived to enable co-estimation of the refractive index of media (e.g. water) in a Visual-Inertial state estimator. The model applies to general cameras housed in a flat port waterproof casing. It is assumed that the refractive interface, i.e. the glass plate is thin (e.g. $2.5\,\mathrm{mm}$) and the camera lens is negligibly close to the interface ($\sim 1\,\mathrm{mm}$).

A. Notations

Below are the notations used in the following sections:

- Boldface letters denote matrices or vectors.
- $\tilde{(.)}$ denotes the unit vector of the given vector.
- $\bar{(.)}$ denotes the normalized vector to a point, normalized by the last dimension of the vector.
- $\mathcal{C}$: the camera coordinate frame.
- $\mathcal{B}$: the IMU (body) coordinate frame.
- $\mathcal{W}$: the world (inertial) coordinate frame.
- $o_{\mathcal{C}}$: the optical axis
- $()_{\diamond,r}$: r in the last subscript denotes that the entity is related to the refractive distortion.
- $()_{\diamond,l}$: l in the last subscript denotes that the entity is related to the lens distortion.

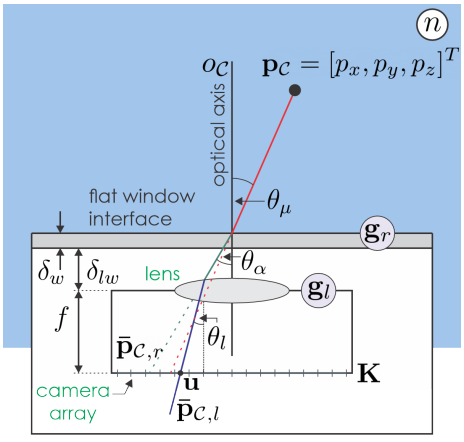

Fig. 2. Visualization of a ray from a point $\mathbf{p}_{\mathcal{C}}$ in the refractive media (e.g. water), undergoing refractive distortion modelled by $\mathbf{g}_r$, followed by lens distortion modelled by $\mathbf{g}_l$. Lastly, the point is projected onto camera array, modelled by $\mathbf{K}$, at pixel coordinate $\mathbf{u}$. The thickness of the refractive interface δ_w and the distance of the interface from the camera lens δ_{lw} are assumed to be small.

B. Refractive distortion

The incidence ray from a point $\mathbf{p}_{\mathcal{C}} = [p_x, p_y, p_z]^T \in \mathbb{R}^3$ in a medium (with refractive index n), expressed in the camera

coordinate frame $\mathcal{C}$, makes an angle θ_μ from the camera optical axis $o_\mathcal{C} = [0, 0, 1]^T$. This ray undergoes refraction upon entering the camera case through the flat-port and is deflected, making an angle θ_α from $o_\mathcal{C}$ after refraction (Figure 2). These angles are related by the Snell's law as

$$\sin\theta_\alpha = n \sin\theta_\mu \tag{1}$$

We write the projection of the point on a plane at a unit distance as $\bar{\mathbf{p}}_\mathcal{C} = \mathbf{g}_p(\mathbf{p}_\mathcal{C})$. It relates $\mathbf{p}_\mathcal{C}$ to a projected coordinate frame such that

$$\bar{\mathbf{p}}_\mathcal{C} = [\bar{p}_x, \bar{p}_y]^T = \left[\frac{p_x}{p_z}, \frac{p_y}{p_z}\right]^T \tag{2}$$

Let $r = \sqrt{p_x^2 + p_y^2}$, then $\sin\theta_\mu$ can be expressed as

$$\sin\theta_\mu = \frac{r}{\sqrt{1 + r^2}} \tag{3}$$

Likewise, we can write for the refracted ray with normalized coordinates as $\bar{\mathbf{p}}_{\mathcal{C},r} = [\bar{p}_{x,r}, \bar{p}_{y,r}]^T$ expressing $\sin\theta_\alpha$

$$\sin\theta_\alpha = \frac{r_r}{\sqrt{1 + r_r^2}} \tag{4}$$

Using equation (1) we can write

$$\frac{r_r}{\sqrt{1 + r_r^2}} = n\frac{r}{\sqrt{1 + r^2}} \tag{5}$$

Let $m = r_r/r$ then by rearranging Eq. (5) we can write m in terms of n and r as

$$m(n, r) = \frac{n}{\sqrt{1 + r^2 - n^2 r^2}} \tag{6}$$

and also in terms of n and r_r as

$$m(n, r_r) = \sqrt{n^2 r_r^2 + n^2 - r_r^2} \tag{7}$$

This allows us to derive the forward (distortion) and inverse (undistortion) maps.

1) Forward Mapping: Accordingly, the forward (distortion) mapping $\bar{\mathbf{p}}_\mathcal{C} \rightarrow \bar{\mathbf{p}}_{\mathcal{C},r}$ takes the form

$$\bar{\mathbf{p}}_{\mathcal{C},r} = m(n, r)\bar{\mathbf{p}}_\mathcal{C} \tag{8a}$$

$$r = \sqrt{\bar{p}_x^2 + \bar{p}_y^2} \tag{8b}$$

2) Inverse Mapping: Similarly, the inverse (undistortion) mapping can be written as $\bar{\mathbf{p}}_{\mathcal{C},r} \rightarrow \bar{\mathbf{p}}_\mathcal{C}$

$$\bar{\mathbf{p}}_\mathcal{C} = \frac{\bar{\mathbf{p}}_{\mathcal{C},r}}{m(n, r_r)} \tag{9a}$$

$$r_r = \sqrt{\bar{p}_{x,r}^2 + \bar{p}_{y,r}^2} \tag{9b}$$

C. Derivations for Iterative Estimation

For integration into VIO, we first need to model the projection of a point $\mathbf{p}_\mathcal{C} \in \mathbb{R}^3$ onto the pixel coordinates $\mathbf{u} \in \mathbb{R}^2$ by incorporating both refractive and lens distortions. A combined model for this can be written as

$$\mathbf{u} = \mathbf{K}\mathbf{g}_l(\mathbf{g}_r(n, \mathbf{g}_p(\mathbf{p}_\mathcal{C}))) \tag{10}$$

where, $\mathbf{K}$ is the camera intrinsics matrix, consisting of focal lengths (f_x, f_y) and image center (c_x, c_y), represented as

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{11}$$

$\mathbf{g}_l$ is the lens distortion function, $\mathbf{g}_r$ is the refractive distortion function, and $\mathbf{g}_p$ is the function that models the projection on a plane at a unit distance. For brevity in the partial derivatives, we drop the function arguments. The partial derivatives of $\mathbf{u}$ (Eq. (10)) w.r.t. (with respect to) $\mathbf{p}_\mathcal{C}$ and n are given by:

$$\frac{\partial \mathbf{u}}{\partial \mathbf{p}_\mathcal{C}} = \mathbf{K}_f \frac{\partial \mathbf{g}_l}{\partial \bar{\mathbf{p}}_{\mathcal{C},r}} \frac{\partial \mathbf{g}_r}{\partial \bar{\mathbf{p}}_\mathcal{C}} \frac{\partial \mathbf{g}_p}{\partial \mathbf{p}_\mathcal{C}} \tag{12a}$$

$$\frac{\partial \mathbf{u}}{\partial n} = \mathbf{K}_f \frac{\partial \mathbf{g}_l}{\partial \bar{\mathbf{p}}_{\mathcal{C},r}} \frac{\partial \mathbf{g}_r}{\partial n} \tag{12b}$$

where $\mathbf{K}_f = diag\,[f_x\ f_y]$ (diagonal matrix). The terms of chain-rule (partial derivatives) along with their function definitions are described below, which are used later in an Iterated Extended Kalman Filter (IEKF) for VIO as in [40].

1) Projection Function: $\bar{\mathbf{p}}_\mathcal{C} = \mathbf{g}_p(\mathbf{p}_\mathcal{C})$

The Jacobian w.r.t. $\mathbf{p}_\mathcal{C}$ takes the form

$$\frac{\partial \mathbf{g}_p}{\partial \mathbf{p}_\mathcal{C}} = \begin{bmatrix} \frac{1}{p_z} & 0 & \frac{-p_x}{p_z^2} \\ 0 & \frac{1}{p_z} & \frac{-p_y}{p_z^2} \end{bmatrix} \tag{13}$$

2) Refractive Distortion Function: $\bar{\mathbf{p}}_{\mathcal{C},r} = \mathbf{g}_r(n, \bar{\mathbf{p}}_\mathcal{C})$

$$\bar{\mathbf{p}}_{\mathcal{C},r} = m\left(n, \sqrt{\bar{p}_x^2 + \bar{p}_y^2}\right)\bar{\mathbf{p}}_\mathcal{C} \tag{14}$$

The Jacobian w.r.t. n takes the form

$$\frac{\partial \mathbf{g}_r}{\partial n} = \begin{bmatrix} \frac{\bar{p}_x\left(\sqrt{h}n^2r^2 + \sqrt{h}^3\right)}{h^2} \\ \\ \frac{\bar{p}_y\left(\sqrt{h}n^2r^2 + \sqrt{h}^3\right)}{h^2} \end{bmatrix} \tag{15}$$

where $h = \left(1 + r^2 - n^2r^2\right)$.

The Jacobian w.r.t. $\bar{\mathbf{p}}_\mathcal{C}$ takes the form

$$\frac{\partial \mathbf{g}_r}{\partial \bar{\mathbf{p}}_\mathcal{C}} = \begin{bmatrix} \frac{n\left(\sqrt{h}\bar{p}_x^2\left(n^2-1\right) + \sqrt{h}^3\right)}{h^2} & \frac{n\bar{p}_x\bar{p}_y\left(n^2-1\right)}{\sqrt{h}^3} \\ \\ \frac{n\bar{p}_x\bar{p}_y\left(n^2-1\right)}{\sqrt{h}^3} & \frac{n\left(\sqrt{h}\bar{p}_y^2\left(n^2-1\right) + \sqrt{h}^3\right)}{h^2} \end{bmatrix} \tag{16}$$

3) Lens Distortion Function: $\bar{\mathbf{p}}_{\mathcal{C},l} = \mathbf{g}_l(\bar{\mathbf{p}}_{\mathcal{C},r})$

In this work we use the equidistant model [41] which is defined as

$$\bar{\mathbf{p}}_{\mathcal{C},l} = \frac{\theta_e}{r_r}\bar{\mathbf{p}}_{\mathcal{C},r} \tag{17a}$$

$$\theta_e = \theta(1 + k_1\theta^2 + k_2\theta^4 + k_3\theta^6 + k_4\theta^8) \tag{17b}$$

$$\theta = \tan^{-1}(r_r) \tag{17c}$$

where r_r is same as in Eq. (9), k_1, k_2, k_3, k_4 are the equidistant model parameters and $\theta_l = \tan^{-1}(\theta_e)$ in Figure 2.

The Jacobians are then given as

$$\frac{\partial \mathbf{g}_l}{\partial \bar{\mathbf{p}}_{\mathcal{C},r}} = \frac{\partial \bar{\mathbf{p}}_{\mathcal{C},l}}{\partial \bar{\mathbf{p}}_{\mathcal{C},r}} + \frac{\partial \bar{\mathbf{p}}_{\mathcal{C},l}}{\partial r_r}\frac{\partial r_r}{\partial \bar{\mathbf{p}}_{\mathcal{C},r}} + \frac{\partial \bar{\mathbf{p}}_{\mathcal{C},l}}{\partial \theta_e}\frac{\partial \theta_e}{\partial \theta}\frac{\partial \theta}{\partial r_r}\frac{\partial r_r}{\partial \bar{\mathbf{p}}_{\mathcal{C},r}} \tag{18}$$

where

$$\frac{\partial \bar{\mathbf{p}}_{\mathcal{C},l}}{\partial \bar{\mathbf{p}}_{\mathcal{C},r}} = diag\begin{bmatrix}\frac{\theta_e}{r_r} & \frac{\theta_e}{r_r}\end{bmatrix} \tag{19a}$$

$$\frac{\partial \bar{\mathbf{p}}_{\mathcal{C},l}}{\partial r_r} = -\frac{\theta_e}{r_r^2}\bar{\mathbf{p}}_{\mathcal{C},r} \tag{19b}$$

$$\frac{\partial r_r}{\partial \bar{\mathbf{p}}_{\mathcal{C},r}} = \frac{1}{r_r}\bar{\mathbf{p}}_{\mathcal{C},r}^T \tag{19c}$$

$$\frac{\partial \bar{\mathbf{p}}_{\mathcal{C},l}}{\partial \theta_e} = \frac{1}{r_r}\bar{\mathbf{p}}_{\mathcal{C},r} \tag{19d}$$

$$\frac{\partial \theta_e}{\partial \theta} = 1 + 3k_1\theta^2 + 5k_2\theta^4 + 7k_3\theta^6 + 9k_4\theta^8 \tag{19e}$$

$$\frac{\partial \theta}{\partial r_r} = \frac{1}{r_r^2 + 1} \tag{19f}$$

The above derivations allow us to use an iterative framework for estimation of the refractive index and to project the points in the world to pixel coordinates and vice-versa. These derivations are not specific to a particular Visual-Inertial Odometry framework and can be applied to a suitable method.

IV. Refractive Visual-Inertial Odometry

The novel refractive camera model tailored to online estimation was integrated into a state-of-the-art VIO system, namely ROVIO [40], to develop a resilient solution for underwater localization in diverse media without any assumption of knowledge of the exact refractive index or need to laboriously calibrate the camera/IMU set-up underwater. The choice of ROVIO was motivated by its robust formulation, delay-free initialization, and good low-light performance as evaluated in [42–44]. It uses multi-level image patches as a frontend and provides the constraints based on photometric error. Its backend is an Iterated Extended Kalman Filter (IEKF). ROVIO uses patchwise QR-decomposition, reducing the error dimensionality and ensuring computational efficiency in the IEKF update step.

We incorporate the estimation of refractive index n in the direct image intensity errors and accordingly re-derive the filter innovation term. ROVIO employs a robocentric formulation thus estimating landmarks relative to the pose of the camera. Estimated landmarks are decomposed and expressed in terms of a bearing vector and an inverse depth parametrization. For the filter formulation, we consider $\mathcal{B}$ as the IMU-fixed coordinate frame, $\mathcal{C}$ as the camera-fixed frame, and $\mathcal{W}$ as the world (inertial) frame, while the resulting state vector and associated covariance are denoted as $\mathbf{s}$ and $\mathbf{\Sigma}$, respectively. The method can support multi-camera systems, but analogous to the original work, here we focus on the general formulation for the monocular case. Note that $\tilde{\mathbf{p}}_{\mathcal{C}}$ denotes the unit vector to the point $\mathbf{p}_{\mathcal{C}}$, following the representation of 3D unit vectors in [40]. The state vector augmented with the refractive index n is then given by

$$\mathbf{s} = [\mathbf{r}\ \mathbf{q}\ \mathbf{v}\ \mathbf{b}_f\ \mathbf{b}_\omega\ \mathbf{c}\ \mathbf{z}\ |n|\ \tilde{\mathbf{p}}_{\mathcal{C},0}, ..., \tilde{\mathbf{p}}_{\mathcal{C},J}\ \rho_0, ..., \rho_J] \tag{20}$$

where $\mathbf{r}, \mathbf{v}$ denote the robocentric position and velocity of the IMU expressed in $\mathcal{B}$, $\mathbf{q}$ is the IMU attitude (map from $\mathcal{B} \rightarrow \mathcal{W}$), $\mathbf{b}_f, \mathbf{b}_\omega$ are the bias of accelerometer and gyroscope expressed in $\mathcal{B}$, $\mathbf{c}, \mathbf{z}$ denote the translational and rotational components of the camera extrinsics against the IMU (maps from $\mathcal{B} \rightarrow \mathcal{C}$), n is modeled as $\dot{n} = \boldsymbol{\mathcal{M}_n}$, where $\boldsymbol{\mathcal{M}_n}$ is the process noise, $\tilde{\mathbf{p}}_{\mathcal{C},j}$ is the bearing vector to the j-th feature (out of maximum J features) expressed in $\mathcal{C}$ and ρ_j is the associated depth parameter under the parametrization $d(\rho_j) = 1/\rho_j$ for the feature distance d_j. Although ROVIO can support extrinsic estimation, in the results presented in this work the extrinsics $\mathbf{c}, \mathbf{z}$ are constant and set to the values estimated offline through Camera-IMU calibration in air. The interested reader may refer to ROVIO [40] for a more detailed formulation of it.

A. Projection model and linear Warping

Given known camera calibration in air, $\mathbf{p}_{\mathcal{C}}$ can be mapped to $\mathbf{u}$ for some n using Eq. (10), denoted as $\mathbf{u} = \pi(n, \tilde{\mathbf{p}}_{\mathcal{C}})$. Also, given the inverse of refractive distortion Eq. (9) and lens distortion, the inverse projection can be written as $\tilde{\mathbf{p}}_{\mathcal{C}} = \pi^{-1}(n, \mathbf{u})$. A linear warping matrix $\boldsymbol{D}$ then accounts for the change in perspective of the patch, while also undergoing distortion in consecutive frames. For two image frames $(1, 2)$ this can be given by stacking the Jacobians of the following functions: inverse projection $\tilde{\mathbf{p}}_{\mathcal{C},1} = \pi^{-1}(n, \mathbf{u}_1)$, process model $\tilde{\mathbf{p}}_{\mathcal{C},2} = f(\tilde{\mathbf{p}}_{\mathcal{C},1})$, and projection $\mathbf{u}_2 = \pi(n, \tilde{\mathbf{p}}_{\mathcal{C},2})$.

Then, the linear warping can be written as

$$\boldsymbol{D} = \frac{\partial \pi(n, \tilde{\mathbf{p}}_{\mathcal{C},2})}{\partial \tilde{\mathbf{p}}_{\mathcal{C},2}}\frac{\partial f(\tilde{\mathbf{p}}_{\mathcal{C},1})}{\partial \tilde{\mathbf{p}}_{\mathcal{C},1}}\frac{\partial \pi^{-1}(n, \mathbf{u}_1)}{\partial \mathbf{u}_1} \in \mathbb{R}^{2\times 2} \tag{21}$$

B. Photometric error

The photometric error of patch j at image pyramid level l $(e_{l,j})$ takes the same form as in ROVIO:

$$e_{l,j}(\mathbf{u}, P, I, \boldsymbol{D}) = P_l(\mathbf{u}_j) - aI_l(\mathbf{u}_l c_l + \boldsymbol{D}\mathbf{u}_j) - b \tag{22}$$

where P is the multi-level image patch centered at $\mathbf{u}$, I is the input image, a and b are the scalars to account for illumination variation in consecutive images, $l \in \{0, \cdots, L\}$ is the patch level and entities with l in subscript denote the computation at that level of the image pyramid, while $c_l = 0.5^l$ is the factor to scale the error based on the pyramid level. The linearized error equation at patch location estimate $\hat{\mathbf{u}}$ can be written as

$$\boldsymbol{e}(\hat{\mathbf{u}} + \delta\mathbf{u}, P, I, \boldsymbol{D}) = \boldsymbol{J}(\hat{\mathbf{u}}, I, \boldsymbol{D})\delta\mathbf{u} + \boldsymbol{e}(\hat{\mathbf{u}}, P, I, \boldsymbol{D}) \quad (23)$$

where $\boldsymbol{e}$ is the stacked errors from Eq. (22), $\boldsymbol{J}$ denotes the corresponding Jacobian and $\delta\mathbf{u}$ is the correction term. Further, the normal equation can be given by

$$\boldsymbol{J}(\hat{\mathbf{u}}, I, \boldsymbol{D})^T\boldsymbol{J}(\hat{\mathbf{u}}, I, \boldsymbol{D})\delta\mathbf{u} = -\boldsymbol{J}(\hat{\mathbf{u}}, I, \boldsymbol{D})^T\boldsymbol{e}(\hat{\mathbf{u}}, P, I, \boldsymbol{D}) \quad (24)$$

To employ dimensionality reduction of the error and its Jacobian, QR decomposition is conducted

$$\boldsymbol{J}(\mathbf{u}_j, I, \boldsymbol{D}_j) = \begin{bmatrix}\boldsymbol{Q}_1(\mathbf{u}_j, I, \boldsymbol{D}) & \boldsymbol{Q}_2(\mathbf{u}_j, I, \boldsymbol{D})\end{bmatrix}\begin{bmatrix}\boldsymbol{R}_1(\mathbf{u}_j, I, \boldsymbol{D}) \\ \mathbf{0}\end{bmatrix} \quad (25)$$

where $\boldsymbol{R}_1(\mathbf{u}_j, I, \boldsymbol{D})$ corresponds to the upper triangular matrix that has full row-rank 2 for distinct corner features, row-rank 1 for line features and $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ have orthogonal columns.

C. Innovation Term

The innovation term $\boldsymbol{y}_{i,j}$ with the projection function incorporating the refractive camera model at i_{th} iteration and j_{th} patch, $\hat{\mathbf{u}} = \pi(n^+, \tilde{\mathbf{p}}^+_{\mathcal{C}\ i,j})$ can be written as

$$\boldsymbol{y}_{i,j} = \boldsymbol{Q}_1(\pi(n^+, \mathbf{p}_{\mathcal{C}^+_{i,j}}), I, \boldsymbol{D}_j)^T\boldsymbol{e}(\pi(n^+, \mathbf{p}_{\mathcal{C}^+_{i,j}}), P_j, I, \boldsymbol{D}_j) \quad (26)$$

where $(\cdot)^+$ is the a-posteriori estimate. Then $\boldsymbol{H}_{i,j}$ can be written as Jacobian of the decomposed innovation term w.r.t $\mathbf{p}_{\mathcal{C}}$ and n respectively as

$$\boldsymbol{H}_{i,j}(\mathbf{p}_{\mathcal{C}}) = \boldsymbol{R}_1(\pi(n^+, \mathbf{p}_{\mathcal{C}^+_{i,j}}), I, \boldsymbol{D}_j)\frac{\partial\pi}{\partial\mathbf{p}_{\mathcal{C}}}(n^+, \mathbf{p}_{\mathcal{C}^+_{i,j}}) \quad (27)$$

$$\boldsymbol{H}_{i,j}(n) = \boldsymbol{R}_1(\pi(n^+, \mathbf{p}_{\mathcal{C}^+_{i,j}}), I, \boldsymbol{D}_j)\frac{\partial\pi}{\partial n}(n^+, \mathbf{p}_{\mathcal{C}^+_{i,j}}) \quad (28)$$

D. Sensitivity Heuristic

Refractive index estimation is vulnerable to sub-pixel errors and noisy tracking of landmarks. Two conditions where the signal-to-noise ratio deteriorates are if the points move along the radial or tangential direction from the image center. Motivated by this, we develop a heuristic map to scale Jacobians which reduces the effect of points near the degraded regions.

Let, $\boldsymbol{T}_{\mathcal{B}_{t+1},\mathcal{B}_t} \in SE(3)$ be the rigid transformation matrix from $\mathcal{B}$ at time t to $\mathcal{B}$ at time $t+1$. Then, the rigid transformation from $\mathcal{C}$ at t to $t+1$ is $\boldsymbol{T}_{\mathcal{C}_{t+1},\mathcal{C}_t} = \boldsymbol{T}_{\mathcal{C}_{t+1},\mathcal{B}_{t+1}}\boldsymbol{T}_{\mathcal{B}_{t+1},\mathcal{B}_t}\boldsymbol{T}_{\mathcal{B}_t,\mathcal{C}_t} \in SE(3)$, where $\boldsymbol{T}_{\mathcal{C}_{t+1},\mathcal{B}_{t+1}} = \boldsymbol{T}^{-1}_{\mathcal{B}_t,\mathcal{C}_t}$ denotes the camera-IMU extrinsics.

Further, assuming a small rotation in frame-to-frame motion, the Epipolar line can approximate the motion of the landmark in image coordinates in consecutive frames. The essential matrix [19] is $\mathbf{E} = [\mathbf{t}]^\times\mathbf{R}$ where $\mathbf{R} \in SO(3)$ is the rotation matrix and $\mathbf{t} \in \mathbb{R}^3$ is the translation from the homogeneous transformation $\boldsymbol{T}_{\mathcal{C}_{t+1},\mathcal{C}_t}$ and the epipolar line in normalized coordinates is $\boldsymbol{\lambda} = \mathbf{E}^T\tilde{\mathbf{p}}_{\mathcal{C},j} \in \mathbb{R}^3$. Let $\tilde{\boldsymbol{\lambda}} \in \mathbb{R}^2$ be the unit vector along the epipolar line in image coordinates, and $\tilde{\mathbf{r}}$ be the unit vector of the line joining the image center and pixel $\mathbf{u}$. Then $\theta = \cos^{-1}(\tilde{\boldsymbol{\lambda}}\cdot\tilde{\mathbf{r}})$ is the angle between the epipolar line and the radial line. The heuristic v can then be given by

$$v = |\sin(2\theta)|^q r^k \quad (29)$$

where $0 < q \leq 1$ and $0 < k \leq 1$ are tunable scalars. We then apply this heuristic patch and level-wise in the VIO estimator by using the scaled Jacobian $v_{i,j}\boldsymbol{H}_{i,j}(n)$ in place of the non-weighted form in Eq. (28). In the following results we use $q = 0.5$ and $k = 0.8$.

V. Evaluation Studies

A host of experiments were conducted to assess the performance of the proposed general refractive camera model and online co-estimation of refractive index and odometry. During the presented studies, we investigate convergence of the estimated refractive index and accompanying odometry results. We initialize the augmented VIO with refractive index values that represent significant changes of media versus the water in which the tests took place. For a general perspective, for liquids at $20°$C, water has a refractive index of 1.33 (wavelength of 587.6nm) [45], acetone's value is 1.36, 60% glucose solution in water has 1.44 [46] and benzene has 1.5. Saline water also assumes values higher than 1.33, e.g. 1.35 for NaCl mass fraction $wt\%$ of 10, wavelength of 589nm and under certain temperatures. Water approaching $100°$C reaches a refractive index of almost 1.31.

A. Robot Experimental Set-up

To verify the proposed contributions, we use an underwater robot (Figure 1) described in [18] that is based on the BlueROV remotely operated vehicle and integrates an Alphasense Core Research Development kit and NVIDIA Orin AGX. The results employ its left forward facing camera with a Field of View (FOV) $D \times H \times V = 165.4° \times 126° \times 92.4°$ and a focal length of 2.4mm measured in air, alongside the time-synchronized BMI085 IMU. The camera is mounted on the top of the robot and is inclined down by $16°$.

B. Dataset for Refractive Visual-Inertial Odometry

The aforementioned robotic system was deployed inside the Marine Cybernetics laboratory (MC-lab) of NTNU which offers a water tank (dimensions: $L \times B \times D = 40\text{m} \times 6.45\text{m} \times 1.5\text{m}$) that is partially supported with Motion Capture (MoCap). Specifically, we utilize the MoCap capabilities above the water by installing a pole and markers on the robot that allow it to be tracked. The MoCap is based on a Qualisys solution and offers partial coverage due to structural occlusions. During the conducted trajectories, all the 5 cameras and IMU of the Alphasense Core Research Development kit are recorded, but as described only the left forward-facing camera is employed for this work. Nevertheless, the complete dataset is released to the community by augmenting our previous release found at `https://github.com/ntnu-arl/underwater-datasets`.

In further detail, a total of 6 trajectories were first conducted and organized based on 3 motion patterns and 2

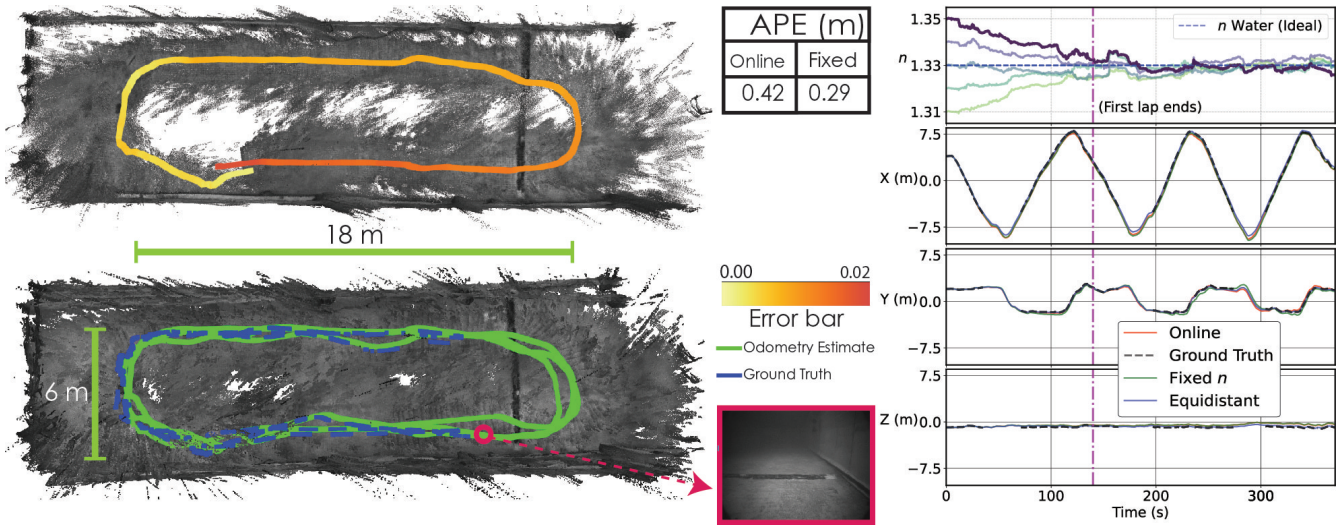

Fig. 3. Detailed evaluation on a rectangular trajectory with good ambient light conditions (Trajectory 1). Top Left: The first lap of the estimated trajectory, given refractive index n initialization of 1.35, with path colorized based on the refractive index absolute error from the ideal value of the refractive index of water $n = 1.33$. The bottom left plot shows the full trajectory for 370 sec along with the ground truth (dashed blue). Top right: the plot showing refractive index n vs. time with initialization of n varying from 1.31 to 1.35. Bottom right: The comparison of odometry against ground truth vs. time for online estimation with initialization from n equal to 1.35, fixed value of n at 1.33, and calibration of the camera directly inside the water of the same pool using an equidistant model is water. The map from the accumulated point cloud is generated using [47] for visualization.

ambient light condition, namely either with ceiling lights on or off. Additionally, one more trajectory was collected to demonstrate the convergence of refractive index is tested under "wild" initial guess. Every mission begins with the robot being approximately at the center of the tank (along the length) and covers of three laps, while the camera always points in the direction of motion. In the trajectories of the first motion pattern, the robot follows a rectangular shape. In the trajectories of the second motion pattern, the robot is piloted in a figure-8 pattern. In the third motion pattern, the robot moves along the length of the path while primarily keeping close the center width-wise. The motivation behind the motion patterns is driven by: i) gradually increasing the distance to visual surfaces seen by the robot, and ii) gradually introducing more variations in the motion. Thus, the rectangular motion patterns allow close proximity to walls, and more reliable visual features are tracked. The figure 8 and the last motion patterns further increase the average distance to visual features and introduce more varying motions.

C. Detailed Evaluation on Single Trajectory

We present the trajectory 1, meaning a rectangular pattern in good ambient light conditions, in detail to highlight the convergence of the refractive index estimated by the proposed method. The results in Figure 3 show the trajectory with a colorized path. We compare it against the odometry estimated with the same camera model initialized and kept fixed at $n = 1.33$, alongside the odometry results using an equidistant camera model calibrated inside the water of this exact pool. As shown, our method enables refractive index estimation and robust odometry without the need of calibration inside the water and despite the initialization from significantly different refractive index value.

D. Collective Evaluation Studies

In this evaluation, we plot the refractive index vs. time plots for all collected trajectories to highlight the convergence of the refractive index given perturbed initial conditions that range from 1.31 to 1.35. Figure 4 presents the associated results of odometry for the first loop of the motion (given initial n of 1.35) alongside the mission-complete plots of n convergence for multiple initial conditions. Figure 5 presents odometry for each mission in full (given initial n of 1.35). The choice of perturbation for the refractive index aims to assess the convergence of refractive index values that cover the range of what water exhibits in very different environmental conditions, thus demonstrating the ability to use the method with no specific knowledge of the medium's refractive index. For trajectories $1 - 3$, good visibility allows the refractive index to converge within ± 0.005 within 150 sec (approximately one loop of the conducted paths). For the low-light trajectories, the convergence is delayed due to a lack of abundant visual features. The lack of visual texture is also highlighted by the sparseness in the point cloud when compared to trajectories with good visibility. Table I summarizes the Absolute Position Error (APE) for the proposed method against ground truth considering a set of significant changes in the refractive index n, alongside the APE for a fixed model where the refractive terms are all calculated for $n = 1.33$ as expected in such waters. Naturally, trajectories with good visibility lead to faster convergence, yet equally importantly the method converges even at low visibility conditions.

E. Refractive Index Converge and Odometry "in the Wild"

Last but not least, we perform an experiment aiming to verify the ability of the proposed estimation solution to converge to the correct refractive index even after extremely wrong initialization. Two tests are conducted, namely with

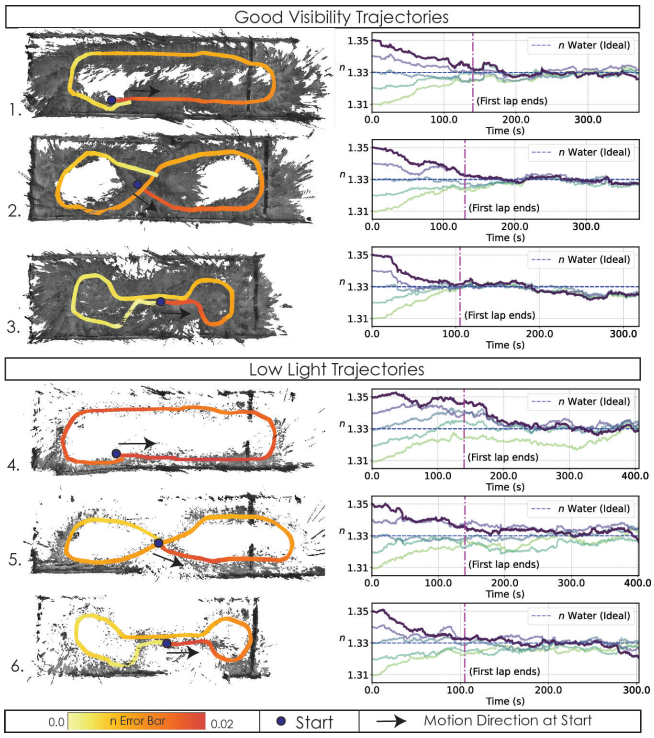

Fig. 4. The top-down plots of trajectory on left show the colorized path based on the absolute difference from the ideal value of the refractive index for water $n = 1.33$ when initialized with 1.35 (purple line in the right plot starting from 1.35). The path is colored according to the error bar shown at the bottom. The blue circle is the starting point of the robot. For clarity of visualization, only the first lap of motion is shown. The refractive index vs. time plots on the right show, for the complete mission, the convergence of refractive index n from perturbations ranging from 1.31 to 1.35. The lack of visual texture results in the sparseness of the point cloud generated using [47] for low-light conditions.

refractive index initially set to 1 (air in Standard Temperature and Pressure (STP)) and 1.6 (a value that for example is assumed by carbon disulfide at 20°C). The robot performs a trajectory involving a figure-8 maneuver. The results regarding the convergence of the refractive index are shown in Figure 6 and demonstrate the resilience of n estimation.

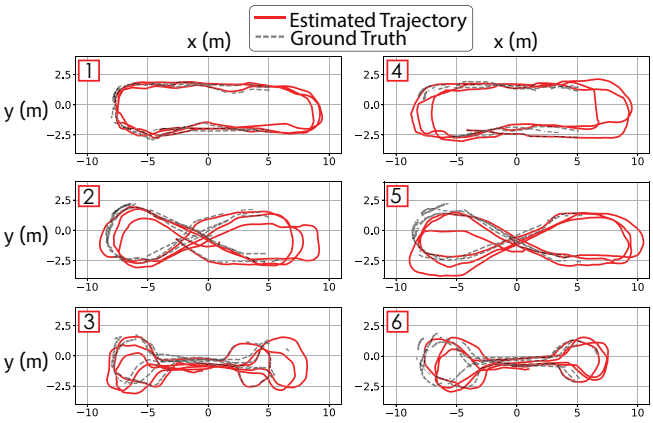

Fig. 5. Collective plots for trajectories comparing against ground truth for an initial n of 1.35. On the left, odometry estimates for trajectories 1-3 (good visibility). Similarly, the plots on the right show the estimates in the low-light trajectories 4-6.

TABLE I

APE(METERS) FOR ESTIMATED ODOMETRY VS. GROUND TRUTH

A. From t=0 (Complete trajectory)						
	Online - Initial n					**Fixed**
no.	**1.31**	**1.32**	**1.33**	**1.34**	**1.35**	**1.33**
1	0.6035	0.4419	0.4337	0.4936	0.4273	0.2962
2	0.6363	0.5411	0.6413	0.7047	0.7332	0.4729
3	0.6306	0.61	0.6022	0.578	0.609	0.5488
4	0.6589	0.5483	0.5488	0.3963	0.5508	0.8611
5	1.2654	1.1362	0.9941	0.9133	0.8226	1.0352
6	0.5341	0.5312	0.6529	0.7791	0.8713	0.446

B. From t=150						
	Online - Initial n					**Fixed**
no.	**1.31**	**1.32**	**1.33**	**1.34**	**1.35**	**1.33**
1	0.4112	0.3997	0.378	0.3701	0.3347	0.2947
2	0.608	0.5782	0.6018	0.5983	0.625	0.4885
3	0.4652	0.4264	0.4356	0.4465	0.4212	0.4504
4	0.5082	0.57	0.4875	0.5352	0.4725	0.5008
5	0.6171	0.7757	0.7669	0.7058	0.8364	1.0026
6	0.3125	0.2776	0.2917	0.2254	0.2794	0.2794

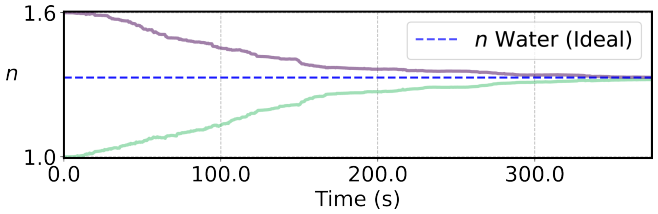

Fig. 6. Refractive index estimation convergence subject to "wild" initialization from n equal to 1 (i.e., that of air at STP) and n equal to 1.6 (e.g., value of carbon disulfide at 20°C).

VI. CONCLUSIONS

This work presented a new general refractive camera model combined with an augmented visual-inertial fusion framework enabling co-estimation of refractive index and odometry. Verified through extensive experimental results, the method allows robust convergence of the refractive index even without a good initialization. Accordingly, it facilitates reliable visual-inertial odometry by only requiring conventional camera/IMU calibration in the air thus eliminating the need for the laborious task of medium-specific calibration or advanced knowledge of a medium's refractive index.

REFERENCES

[1] F. Schill, A. Bahr, and A. Martinoli, "Vertex: A new distributed underwater robotic platform for environmental monitoring," in *Distributed Autonomous Robotic Systems: The 13th International Symposium*. Springer, 2018, pp. 679–693.

[2] E. Galceran and M. Carreras, "Planning coverage paths on bathymetric maps for in-detail inspection of the ocean floor," in *2013 IEEE International Conference on Robotics and Automation*. IEEE, 2013, pp. 4159–4164.

[3] G. A. Hollinger, B. Englot, F. S. Hover, U. Mitra, and G. S. Sukhatme, "Active planning for underwater inspection and the benefit of adaptivity," *The International Journal of Robotics Research*, vol. 32, no. 1, pp. 3–18, 2013.

[4] Y. Wu, X. Ta, R. Xiao, Y. Wei, D. An, and D. Li, "Survey of underwater robot positioning navigation," *Applied Ocean Research*, vol. 90, p. 101845, 2019.

[5] A. Bahr, J. J. Leonard, and M. F. Fallon, "Cooperative localization for autonomous underwater vehicles," *The International Journal of Robotics Research*, vol. 28, no. 6, pp. 714–728, 2009.

[6] L. Paull, S. Saeedi, M. Seto, and H. Li, "Auv navigation and localization: A review," *IEEE Journal of oceanic engineering*, vol. 39, no. 1, pp. 131–149, 2013.

[7] Y. Xu, R. Zheng, S. Zhang, and M. Liu, "Robust inertial-aided underwater localization based on imaging sonar keyframes," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022.

[8] H. Johannsson, M. Kaess, B. Englot, F. Hover, and J. Leonard, "Imaging sonar-aided navigation for autonomous underwater harbor surveillance," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2010, pp. 4396–4403.

[9] A. Shukla and H. Karki, "Application of robotics in offshore oil and gas industry—a review part ii," *Robotics and Autonomous Systems*, vol. 75, pp. 508–524, 2016.

[10] M. Ferrera, V. Creuze, J. Moras, and P. Trouvé-Peloux, "Aqualoc: An underwater dataset for visual–inertial–pressure localization," *The International Journal of Robotics Research*, vol. 38, no. 14, pp. 1549–1559, 2019.

[11] R. Miao, J. Qian, Y. Song, R. Ying, and P. Liu, "Univio: Unified direct and feature-based underwater stereo visual-inertial odometry," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–14, 2021.

[12] B. Teixeira, H. Silva, A. Matos, and E. Silva, "Deep learning for underwater visual odometry estimation," *IEEE Access*, vol. 8, pp. 44 687–44 701, 2020.

[13] S. Rahman, A. Q. Li, and I. Rekleitis, "Svin2: An underwater slam system using sonar, visual, inertial, and depth sensor," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 1861–1868.

[14] Y. Randall, "Flsea: Underwater visual-inertial and stereo-vision forward-looking datasets," Ph.D. dissertation, University of Haifa (Israel), 2023.

[15] B. Joshi, H. Damron, S. Rahman, and I. Rekleitis, "Sm/vio: Robust underwater state estimation switching between model-based and visual inertial odometry," *arXiv preprint arXiv:2304.01988*, 2023.

[16] C. Gu, Y. Cong, and G. Sun, "Environment driven underwater camera-imu calibration for monocular visual-inertial slam," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 2405–2411.

[17] R. W. Austin and G. Halikas, "The index of refraction of seawater," 1976.

[18] M. Singh, M. Dharmadhikari, and K. Alexis, "An online self-calibrating refractive camera model with application to underwater odometry," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 10 005–10 011.

[19] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[20] A. W. Fitzgibbon, "Simultaneous linear estimation of multiple view geometry and lens distortion," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1. IEEE, 2001, pp. I–I.

[21] J. P. Barreto and K. Daniilidis, "Fundamental matrix for cameras with radial distortion," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, vol. 1. IEEE, 2005, pp. 625–632.

[22] R. G. Willson and S. A. Shafer, "What is the center of the image?" *JOSA A*, vol. 11, no. 11, pp. 2946–2955, 1994.

[23] V. Chari and P. Sturm, "Multiple-view geometry of the refractive plane," in *BMVC 2009-20th British machine vision conference*. The British Machine Vision Association (BMVA), 2009, pp. 1–11.

[24] L. Huang, X. Zhao, S. Cai, and Y. Liu, "Plate refractive camera model and its applications," *Journal of Electronic Imaging*, vol. 26, no. 2, pp. 023 020–023 020, 2017.

[25] T. Treibitz, Y. Schechner, C. Kunz, and H. Singh, "Flat refractive geometry," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 1, pp. 51–65, 2011.

[26] A. Sedlazeck and R. Koch, "Perspective and non-perspective camera models in underwater imaging–overview and error analysis," in *Outdoor and Large-Scale Real-World Scene Analysis: 15th International Workshop on Theoretical Foundations of Computer Vision, Dagstuhl Castle, Germany, June 26-July 1, 2011. Revised Selected Papers*. Springer, 2012, pp. 212–242.

[27] A. Agrawal, S. Ramalingam, Y. Taguchi, and V. Chari, "A theory of multi-layer flat refractive geometry," in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 3346–3353.

[28] A. Jordt-Sedlazeck and R. Koch, "Refractive structure-from-motion on underwater images," in *Proceedings of the IEEE international Conference on Computer Vision*, 2013, pp. 57–64.

[29] R. Kawahara, S. Nobuhara, and T. Matsuyama, "A pixel-wise varifocal camera model for efficient forward projection and linear extrinsic calibration of underwater cameras with flat housings," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2013, pp. 819–824.

[30] S. Chaudhury, T. Agarwal, and P. Maheshwari, "Multiple view 3-d reconstruction in water," in *2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*. IEEE, 2015, pp. 1–4.

[31] S. Haner and K. Astrom, "Absolute pose for cameras under flat refractive interfaces," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1428–1436.

[32] X. Hu, F. Lauze, and K. S. Pedersen, "Refractive pose refinement: Generalising the geometric relation between camera and refractive interface," *International Journal of Computer Vision*, vol. 131, no. 6, pp. 1448–1476, 2023.

[33] X. Hu, F. Lauze, K. S. Pedersen, and J. Mélou, "Absolute and relative pose estimation in refractive multi view," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 2569–2578.

[34] P. Zhang, Z. Wu, J. Wang, S. Kong, M. Tan, and J. Yu, "An open-source, fiducial-based, underwater stereo visual-inertial localization method with refraction correction," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 4331–4336.

[35] J. Servos, M. Smart, and S. L. Waslander, "Underwater stereo slam with refraction correction," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 3350–3355.

[36] S. Suresh, E. Westman, and M. Kaess, "Through-water stereo slam with refraction correction for auv localization," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 692–699, 2019.

[37] F. Shkurti, I. Rekleitis, M. Scaccia, and G. Dudek, "State estimation of an underwater robot using visual and inertial information," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2011, pp. 5054–5060.

[38] S. Rahman, A. Q. Li, and I. Rekleitis, "Sonar visual inertial slam of underwater structures," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 5190–5196.

[39] C. Hu, S. Zhu, Y. Liang, and W. Song, "Tightly-coupled visual-inertial-pressure fusion using forward and backward imu preintegration," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 6790–6797, 2022.

[40] M. Bloesch, M. Burri, S. Omari, M. Hutter, and R. Siegwart, "Iterated extended kalman filter based visual-inertial odometry using direct photometric feedback," *The International Journal of Robotics Research*, vol. 36, no. 10, pp. 1053–1072, 2017.

[41] J. Kannala and S. Brandt, "A Generic Camera Model and Calibration Method for Conventional, Wide-Angle, and Fish-Eye Lenses," vol. 28, pp. 1335–40.

[42] M. Tranzatto, T. Miki, M. Dharmadhikari, L. Bernreiter, M. Kulkarni, F. Mascarich, O. Andersson, S. Khattak, M. Hutter, R. Siegwart *et al.*, "Cerberus in the darpa subterranean challenge," *Science Robotics*, vol. 7, no. 66, p. eabp9742, 2022.

[43] M. Tranzatto, M. Dharmadhikari, L. Bernreiter, M. Camurri, S. Khattak, F. Mascarich, P. Pfreundschuh, D. Wisth, S. Zimmermann, M. Kulkarni *et al.*, "Team cerberus wins the darpa subterranean challenge: Technical overview and lessons learned," *arXiv preprint arXiv:2207.04914*, 2022.

[44] M. Tranzatto, F. Mascarich, L. Bernreiter, C. Godinho, M. Camurri, S. Khattak, T. Dang, V. Reijgwart, J. Loeje, D. Wisth *et al.*, "Cerberus: Autonomous legged and aerial robotic exploration in the tunnel and urban circuits of the darpa subterranean challenge," *arXiv preprint arXiv:2201.07067*, 2022.

[45] M. Daimon and A. Masumura, "Measurement of the refractive index of distilled water from the near-infrared region to the ultraviolet region," *Applied optics*, vol. 46, no. 18, pp. 3811–3820, 2007.

[46] D. R. Lide, *CRC handbook of chemistry and physics*. CRC press, 2004, vol. 85.

[47] L. Lipson, Z. Teed, and J. Deng, "Raft-stereo: Multilevel recurrent field transforms for stereo matching," in *International Conference on 3D Vision (3DV)*, 2021.