

Underwater Dome-Port Camera Calibration: Modeling of Refraction and Offset through N-Sphere Camera Model

Monika Roznere¹, Adithya K. Pediredla¹, Samuel E. Lensgraf¹, Yogesh Girdhar², and Alberto Quattrini Li¹

Abstract—The optical effects that are observed in underwater imagery are more complex than those in-air. This is partially because we enclose most underwater cameras in a watertight enclosure, such as a hemispheric dome window. We then observe optical issues including the distortion effects of the lens, e.g., wide-angle field-of-view (FOV), the refractive effects at the enclosure (water-acrylic and acrylic-air) interfaces, and offset effects of a non-centered camera with respect to the dome. In this paper, we present an N-Sphere (NS) and Shifted N-Sphere (S-NS) camera models, tailored to these cameras and lenses mounted in water-tight dome enclosures. The proposed camera models treat each layer of effects as a ‘sphere’ that a 3D point will project on. Furthermore, the S-NS model includes additional parameters to address the camera offset variability. The versatility of the NS model makes it applicable to various lenses, as validated with fisheye (FOV > 120°) and wide-FOV (FOV ≈ 120°). We validated our models with different in-water calibration sequences, lenses, and housing setups, as well as with comparisons with other state-of-the-art camera models. Additionally, we demonstrated the performance of our proposed models in an example stereo-based visual odometry application. The low computational load of the proposed models makes it ideal for integrating in real-time visual navigation and reconstruction frameworks. We provide full math derivations of the proposed models as well as example C++ header files¹ for easy incorporation in independent projects.

I. INTRODUCTION

In this paper, we present the N-Sphere (NS) and Shifted N-Sphere (S-NS) camera models, tailored to cameras mounted in water-tight hemispheric dome enclosures. Our models are able to address the image distortions, as illustrated in Fig. 1, that are due to the medium change at the enclosure’s interface and the off-centering of the camera in relation to the dome. In addition, the models are applicable to a variety of camera lenses, including those with wide-angle field-of-view (FOV). It is computationally light in its projection operations, which is vital for the techniques used in robot visual navigation and 3D reconstruction. As such, an improved camera model will help elevate the work in applications like coral reef and seabed monitoring [1], [2], construction inspections [3], [4], and archaeological site assessment [5], [6].

With the advancement in underwater sensor and robot systems, especially with the use of Autonomous Underwater Vehicles (AUVs), there is an increased interest in conducting

¹Dartmouth College, Hanover, NH, USA, 03755
{monika.roznere.gr, adithya.k.pediredla,
samuel.e.lensgraf.gr, alberto.quattrini.li}
@dartmouth.edu

²Woods Hole Oceanographic Institution (WHOI), Falmouth, MA, USA,
02543 ygirdhar@whoi.edu

¹<https://github.com/dartmouthrobotics/shifted-nsphere-camera-model>

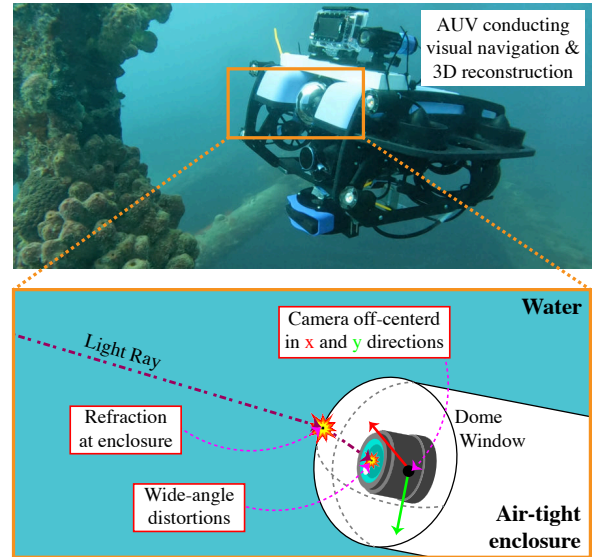


Fig. 1: Illustration of an Autonomous Underwater Vehicle (AUV) in the midst of conducting vision-based navigation and 3D reconstruction. The images captured by the AUV appear distorted due to water-glass-air medium changes at the enclosure, wide-angle lens-based refraction, and camera off centering with respect to the dome. Our proposed camera models address these issues to help aid in underwater robotic operations.

complex aquatic missions, such as cataloging reefscape [7], tracking animals [8], inspecting marine infrastructure [9], and exploring seafloor terrains [10], [11]. A commonality across these different studies is the necessity for extracting high resolution data that are geometrically accurate. While sonars provide accurate geometric descriptions, only cameras are capable of capturing scene color and texture. Therefore, there is a lot of work in designing and implementing vision-based algorithms, from Photometric Stereo [12] and Structure-from-Motion [13] to Visual Odometry (VO) [14] and Visual Simultaneous Localization and Mapping (VSLAM) [15], to aid in these underwater operations. However, various degradation effects, such as light absorption and scattering during light travel in water, affect camera imagery [16], [17]. Two other challenges, which are the focus of this paper, are the distortion effect due to the medium change at the camera housing interface [18] and the offset-based effect due to the camera’s off-centering with respect to the dome [19]. If left unresolved, these issues will negatively affect the geometric accuracy of the applied vision-based algorithms [20].

To utilize a camera underwater, a water-tight housing is necessary. But, during a light ray’s travel to the camera image sensor, it will change directions twice given the two medium transitions, from water to housing (typically acrylic-based)

and housing to air. The significance in the change of direction depends on the two mediums' refraction indexes and the ray's incident angle with respect to the housing's surface normal. In addition, sea water's index of refraction is the most irregular due to its contingency on pressure, temperature, and salinity [21]. Since the effect of the refraction index is possible to estimate [22], most underwater camera calibration work focuses on how the camera enclosure's shape affects the image distortion [18], [19], [23], [24].

There are two common underwater camera housings: flat-pane windows and hemispheric domes. While flat-pane windows are attractive for their simplicity in built, they induce challenging distortion effects that make it difficult to model, violating the popular pinhole projection model [18]. Moreover, the distortion effect due to the flat-pane window reduces the camera's FOV, making them incompatible with cameras with wide-angle lenses.

Hemispheric dome housings are prevalent in many affordable underwater robots (e.g., BlueROV2). Unlike with flat-pane housings, if the camera's optical center is at the center of the hemispheric dome, then one can assume the pinhole projection model. In other words, all incident light rays will hit the dome glass at its surface normal. Thus, no refraction will occur during its pass into the second medium (air) to the camera sensor. However, this assumption is not always true, especially for stereo cameras under one dome [19], cameras with more complex lenses, such as wide-angle (FOV $\approx 120^\circ$) and fisheye (FOV $> 120^\circ$) [25], or cameras that are unintentionally off-centered during placement. But, with the rise in use of fisheye and wide-angle lenses for improving an AUV's FOV [26], dome housings are necessary to prevent any periphery clippings.

Therefore, this paper focuses on advancing the underwater camera model for cameras of various lens types, like wide-angle and fisheye, placed in hemispheric domes. Specifically, we are motivated by three inspirations. 1) The work on the Double Sphere (DS) camera model by Usenko *et al.* [27] shows it to be geometrically accurate for fisheye cameras, while computationally faster than the state-of-the-art models, like the Kannala-Brandt [28]. 2) The observation by Iscar and Johnson-Roberston [19] that a dome housing acts like an additional lens in the camera-dome optical system. 3) The need for a less time-intensive method for calibrating cameras, not centered within a dome enclosure [19]. Hence, in this paper we show that by extending the DS model to include additional spheres, or lenses (the dome), as well as offset-based parameters, a new model can more accurately represent the distortion effects that occur with wide-angle and fisheye cameras in underwater dome housings.

The structure of the paper is as follows: in Section II, we provide an overview of various camera calibration techniques applied to underwater applications. In Section III, we introduce the proposed N-Sphere and the Shifted N-Sphere camera models and their derivations. In Section IV, we compare the proposed models with the state-of-the-art camera models, showcasing the following contributions:

- The proposed NS and S-NS camera models are apt for

cameras mounted in water-tight dome enclosures.

- The proposed models are applicable to various lenses, including fisheye and wide-FOV lenses.
- The proposed models fit under a set of diverse environments and camera-housing setups, with an analysis of other state-of-the-art models.
- The proposed models preserve the overall computational benefits that DS already displayed and, as demonstrated, are well-suited for integrating in real-time VO/VIO or VSLAM systems.

While not evaluated, we believe the proposed models should be applicable to standard cameras (FOV $< 120^\circ$) in dome-port enclosures. Overall, improving the underwater camera model will strengthen the foundation for robust and accurate underwater state estimation and 3D reconstruction, necessary for a large variety of operations, such as preservation of archaeological sites, surveying reef ecosystems, and inspecting man-made constructions.

II. RELATED WORK

There is extensive work on modeling the underwater image formation and related medium-induced effects. This includes light attenuation due to absorption and scattering, and light refraction due to water-air transitions. Here we focus on another topic of interest, representing the distortion effects due to the shape of the camera enclosure, which is typically either a flat-pane window or hemispheric dome, as well as the positioning of the camera in the enclosure.

Flat-pane windows are a simple, flexible choice, especially for designing water-tight housings. However, the refraction at the pane interface leads to significant image distortions that are challenging to handle. For instance, it reduces the camera's FOV by 25% [6], [29]. More importantly, cameras in flat-pane housings do not possess a single viewpoint, thus are incompatible with the perspective projection model [18].

Łuczyński *et al.* addressed this issue by introducing the Pinax model [23], which adopts both the pinhole model characteristics, assuming that the location of the camera is as close to the window pane as possible, and the axial model's projection function from [30]. While the model works well, it does require assumptions, including the camera's distance from the pane, the pane's thickness, and the refraction index of the water.

A camera in a hemispheric dome port adheres to the perspective projection model as long as the camera's optical center is at the dome's center [25] and the lens in use has limited radial distortion (FOV $< 120^\circ$). With a valid setup, each incoming light ray's angle aligns with the surface normal at its incident point, resulting in no bending during the transition from water to air. This assumption is difficult to comply with since it requires perfectly manufactured domes [25] and precise camera positioning. One approach is to model the light ray's paths during the changes at medium interfaces. The name of this technique is raytracing, and typically requires calibration data above and below water [31]. On the other hand, miniscule offsets only induce negligible errors, easily ignored or absorbed by camera parameters

during calibration [32]. However, many still observe the importance of addressing the refraction effects at the water-acrylic and acrylic-air interfaces of the enclosure [33].

Furthermore, we cannot always center cameras in a dome, for example a stereo camera system positioned under one dome viewport. Under this circumstance, Iscar and Johnson-Roberson [19] were able to recover the camera position relative to the dome by observing the measured distortions and utilizing the point spread function that characterizes the camera-lens-dome system. However, the calibration approach is time intensive and requires prior knowledge of the mediums' indices of refraction as well as the point spread function measurements for possible camera-dome positions. Similarly, in the study by Pedersen *et al.* [34], they observed that action-cameras, such as GoPro and Intova, did not present good 3D reconstruction results metric-wise. It was most likely due to the distortions from the small focal length lenses, but also possibly due to the unsuitable dome enclosures. Since they calibrated the associated data using the pinhole model with radial-tangential distortions, the authors concluded that these cameras require a more suitable model.

Interestingly, during Iscar's and Johnson-Roberson's [19] analysis on cameras housed in domes, they observed space-varying defocus effects. They reflected that a hemispheric dome approximately acts as another lens in the optical system. This remark led to the interest of the work on the Double Sphere (DS) camera model by Usenko *et al.* [27]. They extended the Unified Camera Model (UCM) [35], already designed for catadioptric cameras, such that it could better represent a wide variety of fisheye lenses. Importantly, they showed that the computation load of the projection and unprojection operations is low, compared to other well known fisheye camera models, including Kannala-Brandt [28], which is currently the state-of-the-art method for fisheye camera calibration [10]. It is to our knowledge that no underwater work has utilized the UCM or DS camera models in their calibration procedure.

If the DS model can accurately model the image formation of fisheye or wide angle lenses, then introducing another sphere in the model could then account for the refraction due to the hemispheric dome. Moreover, if there is camera positioning offset from the center of the dome, then this could potentially represent as an offset of a portion of the spheres in the model. Therefore, this paper proposes the N-Sphere (NS) and Shifted N-Sphere (S-NS) camera models to address the need of an underwater camera model for wide-FOV cameras, enclosed in a dome enclosure with potential off-centering.

III. N-SPHERE AND SHIFTED N-SPHERE CAMERA MODELS

Before discussing the proposed camera models, we will briefly define some notations used in the paper. Lowercase letters will denote scalars, e.g., u , while bold lowercase letters will denote vectors, e.g., \mathbf{x} .

We represent pixel coordinates as $\mathbf{u} = [u, v]^T \in \Theta \subset \mathbb{R}^2$, and 3D point coordinates as $\mathbf{x} = [x, y, z]^T \in \Omega \subset \mathbb{R}^3$. Note,

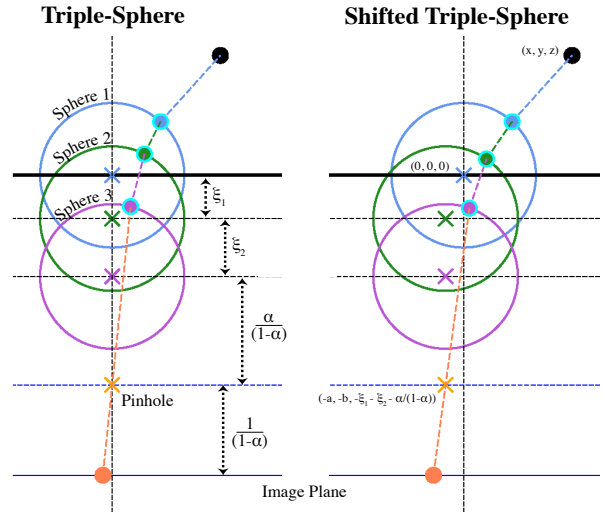


Fig. 2: Schematic illustrations of the N-Sphere (NS) (left) and Shifted N-Sphere (S-NS) (right) camera models for the case of $n = 3$, or so-called Triple-Sphere (TS) and Shifted Triple-Sphere (S-TS) camera models. A 3D point (black) projects onto sphere 1 (blue), located at the origin for both TS and S-TS. The point then projects consecutively onto spheres 2 (green) and 3 (purple), offset by ξ_1 and ξ_2 in the z axis, respectively. Therefore, the center of sphere 3 is located $\xi_1 + \xi_2$ from sphere 1. The pinhole (orange X) is located $\frac{\alpha}{1-\alpha}$ from sphere 3. In the S-TS case, spheres 2 and 3 and the pinhole shift by a and b offsets, in the x and y axes, respectively. The image plane is then located $\frac{1}{1-\alpha}$ from the pinhole.

Θ is the set of valid pixel coordinates on the image plane, while Ω is the set of valid 3D point projections. The camera coordinate frame is as follows. The z axis aligns with the principal axis of the camera, and the x and y axes align with the corresponding axes on the image plane.

Regarding the camera model operations, the projection function is the mapping $\pi : \Omega \rightarrow \Theta$. The inverse or the unprojection function is the mapping $\pi^{-1} : \Theta \rightarrow \Omega$. Note that the result of the unprojection function is a bearing vector of unit length.

A. N-Sphere Camera Model

The proposed N-Sphere Sphere (NS) camera model is an extension of the DS camera model [27], which is itself an extension of the UCM [35]. Here, there are n number of spheres; if $n = 2$ then it would entail the DS model, and if $n = 1$ then it would entail the UCM. Note that when $n = 0$, we result back to the pinhole camera model. The conception of the n spheres is to approximate the different levels of refraction that might occur in the lens/camera level as well as the medium change at the dome enclosure interface. Fig. 2 left illustrates an example of the NS camera model where $n = 3$, or so-called Triple-Sphere (TS) camera model.

In the NS camera model, a point consecutively projects onto the n number of spheres, each offset by the previous sphere i by $\xi_i \in [-1, 1]$ along the z axis. Note, a positive ξ_i moves the sphere $i + 1$ along the negative direction of the z axis. Afterwards, the final point projects onto the image plane via the pinhole, offset by $\frac{\alpha}{1-\alpha}$ from the last sphere, where $\alpha \in [0, 1]$. Furthermore, the image plane is off by $\frac{1}{1-\alpha}$ from

the pinhole plane. Overall, the NS camera model contains $n + 4$ parameters $\mathbf{i} = [f_x, f_y, c_x, c_y, \alpha, \xi_1, \dots, \xi_{n-1}]^T$

The projection function is as follows:

$$\pi(\mathbf{x}, \mathbf{i}) = \begin{bmatrix} f_x \frac{x}{\alpha d_n - (1-\alpha)(z + \sum_{j=1}^{n-1} \xi_j d_j)} \\ f_y \frac{y}{\alpha d_n - (1-\alpha)(z + \sum_{j=1}^{n-1} \xi_j d_j)} \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \end{bmatrix}, \quad (1)$$

$$d_j = \sqrt{x^2 + y^2 + (z + \sum_{q=1}^{j-1} \xi_q d_q)^2}, \quad (2)$$

where projections are valid for:

$$\Omega = \{\mathbf{x} \in \mathbb{R}^3 | z > -w_n d_1\}, \quad (3)$$

$$w_n = \frac{w_0 + \sum_{j=1}^{n-1} \xi_j}{\sqrt{(w_0 + \sum_{j=1}^{n-1} \xi_j)^2 - w_0^2 + 1}}, \quad (4)$$

$$w_0 = \begin{cases} \frac{\alpha}{1-\alpha} & \text{if } \alpha \leq 0.5 \\ \frac{1}{1-\alpha} & \text{if } \alpha > 0.5 \end{cases} \quad (5)$$

The unprojection function is as follows:

$$\pi^{-1}(\mathbf{u}, \mathbf{i}) = k_{\xi_1} \begin{bmatrix} m_x \\ m_y \\ 1 \end{bmatrix} \begin{bmatrix} z_\epsilon \prod_{j=2}^{n-1} z_{\xi_j} \\ z_\epsilon \prod_{j=2}^{n-1} z_{\xi_j} \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ \xi_1 \end{bmatrix}, \quad (6)$$

$$z_\epsilon = \frac{k_\epsilon}{k_\epsilon - \frac{\alpha}{1-\alpha}}, \quad z_{\xi_j} = \frac{k_j}{k_j - \xi_j}, \quad (7)$$

$$k_\epsilon = \frac{\frac{\alpha}{1-\alpha} + \sqrt{1 + (1 - (\frac{\alpha}{1-\alpha})^2)r^2}}{1 + r^2}, \quad (8)$$

$$k_{\xi_j} = \frac{\xi_j \pm \sqrt{1 + (1 - \xi_j^2)r^2 z_\epsilon^2 \prod_{p=j+1}^{n-1} z_{\xi_p}^2}}{1 + r^2 z_\epsilon^2 \prod_{p=j+1}^{n-1} z_{\xi_p}^2}, \quad (9)$$

$$m_x = \frac{u - c_x}{f_x}(1 - \alpha), \quad m_y = \frac{v - c_y}{f_y}(1 - \alpha), \quad (10)$$

$$r^2 = m_x^2 + m_y^2. \quad (11)$$

Here, unprojections are valid for:

$$\Theta = \begin{cases} \mathbb{R}^2 & \text{if } \alpha \leq 0.5 \\ \{\mathbf{u} \in \mathbb{R}^2 | r^2 \leq \frac{1}{2\alpha-1}\} & \text{if } \alpha > 0.5 \end{cases} \quad (12)$$

As one may notice, the NS camera model is represented and can be coded in a dynamic programming fashion.

B. Shifted N-Sphere Camera Model

The proposed Shifted N-Sphere (S-NS) camera model addresses the common issue of centering the camera in a dome enclosure. In many cases, there is some offset along the x and y axes. In the S-NS model, the first sphere is at the origin, but the consecutive spheres are off by a and b with respect to the x and y axes. Consequently, this will move the pinhole as well. Fig. 2 right presents an example of the Shifted Triple-Sphere (S-TS) camera model. The number of parameters for this model increases by two: $\mathbf{i} = [f_x, f_y, c_x, c_y, \alpha, \xi_1, \dots, \xi_{n-1}, a, b]^T$.

Since the S-NS model is nearly identical to the NS model, we only report the equations with modifications (highlighted in red).

The projection function is as follows:

$$\pi(\mathbf{x}, \mathbf{i}) = \begin{bmatrix} f_x \frac{x_\Delta}{\alpha d_n - (1-\alpha)(z + \sum_{j=1}^{n-1} \xi_j d_j)} \\ f_y \frac{y_\Delta}{\alpha d_n - (1-\alpha)(z + \sum_{j=1}^{n-1} \xi_j d_j)} \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \end{bmatrix}, \quad (13)$$

$$d_j = \begin{cases} \sqrt{x^2 + y^2 + z^2} & \text{if } j = 1 \\ \sqrt{x_\Delta^2 + y_\Delta^2 + (z + \sum_{q=1}^{j-1} \xi_q d_q)^2} & \text{o/w} \end{cases} \quad (14)$$

$$x_\Delta = x + ad_1, \quad y_\Delta = y + bd_1, \quad (15)$$

where projections are valid for:

$$\Omega = \{\mathbf{x} \in \mathbb{R}^3 | z > -w_n d_\Delta\}, \quad (16)$$

$$d_\Delta = \sqrt{x_\Delta^2 + y_\Delta^2 + z^2}. \quad (17)$$

The unprojection operation is as follows:

$$\pi^{-1}(\mathbf{u}, \mathbf{i}) = k_{\xi_1} \begin{bmatrix} m_x \\ m_y \\ 1 \end{bmatrix} \begin{bmatrix} z_\epsilon \prod_{j=2}^{n-1} z_{\xi_j} \\ z_\epsilon \prod_{j=2}^{n-1} z_{\xi_j} \\ 1 \end{bmatrix} - \begin{bmatrix} a \\ b \\ \xi_1 \end{bmatrix}, \quad (18)$$

$$k_{\xi_1} = \frac{\xi_{1\Delta} \pm \sqrt{\xi_{1\Delta}^2 - \delta(1 + r^2 z_\epsilon^2 \prod_{p=j+1}^{n-1} z_{\xi_p}^2)}}{1 + r^2 z_\epsilon^2 \prod_{p=j+1}^{n-1} z_{\xi_p}^2}, \quad (19)$$

$$\xi_{1\Delta} = am_x z_\epsilon \prod_{p=j+1}^{n-1} z_{\xi_p} + bm_y z_\epsilon \prod_{p=j+1}^{n-1} z_{\xi_p} + \xi_1, \quad (20)$$

$$\delta = a^2 + b^2 + \xi_1^2 - 1, \quad (21)$$

and Θ still holds as defined in the NS camera model.

C. Camera Model Parameters Optimization

To optimize the camera model parameters, we rely on a target with known size visible from different camera poses, e.g., checkerboard. By detecting corners on the target, the procedure solves an optimization problem, minimizing the reprojection error. We use the Gauss-Newton algorithm to iteratively solve the optimization problem. Note that we calculated the corresponding Jacobians for the projection and unprojection functions (not reported here for space constraints).

IV. EVALUATION

A. Camera Calibration

We evaluated our proposed models with various calibration sequences, collected in-air and in-water settings with different camera lens and dome housing setups.

We used two fisheye lenses: an ADL-HFE1414C [36] with a 1.4 mm focal length and a 185° FOV, and a FOCTek CS-M2.5IR [37] with a 2.5 mm focal length and 190° FOV. For ease, we will shorthand them as ADL (185°) and FOCTek (190°), respectively. We used a BFS-U3-16S2C-CS camera [38] (blackfly's 1440x1080) with a 5 MP resolution, a horizontal FOV of 80°, and a vertical FOV of 64°.

We collected the main calibration sequences in a freshwater river with either a 3 or 4 in diameter dome, utilizing a checkerboard as the calibration target. Additionally, we included sequences captured with the ADL (185°) lens in a swimming pool and in air, with and without the 4 in dome. For each dataset scenario, we created three train-test datasets, where the training dataset consists of 67% of the data and the rest considered for testing.

We also included a calibration sequence captured with one of CUREE AUV’s [39] black and white stereo cameras housed in a 2 in diameter dome. The camera is an OAK-FFC OV9282 M12 [40] with a 1 MP (1280x8000) resolution, a horizontal FOV of 80°, and a vertical FOV of 54°. Its lens is an ArduCam M23325H12 [41] with a 3.25 mm focal length and a 120° FOV. We will note this camera as ArduCam (120°). The corresponding calibration dataset, however, consisted of only 12 images of a 9 x 14 ChArUco board (104 corners). To compensate, we applied the leave-one-out cross-validation method, a special case of the k-fold cross-validation [42]. As such, we created 12 datasets with 11 training images and 1 testing image.

Our proposed models that we chose to evaluate include TS, Quadruple-Sphere (QS), Shifted UCM (S-UCM), Shifted DS (S-DS), and Shifted TS (S-TS). We compared our proposed models with KB [28], commonly used for fisheye lenses, UCM [35], and DS [27]. In addition, we ran the pinhole camera model with radial-tangential distortion (Pinhole RT) on the sequence using the OAK-FFC (120°) camera. We utilized the calibration module within the VIO system [43], called *basalt*², created by the same authors of the DS model, since it has all the camera calibration modules of interest.

For fair comparison, we ran two rounds of optimization. The first round of optimization initialized the model with default values, with many points were likely to be rejected as outliers. After convergence, we used the new intrinsic parameter values for the second round of calibration. At this time, almost all points in the training dataset stay during optimization. If not, we run another round. Finally, we use the converged intrinsic parameters for evaluation.

After optimizing the calibration parameters with the training dataset, we evaluate the models with the testing dataset over an average of ten rounds. Table I reports the mean reprojection error (in pixels) based on the testing dataset, the corresponding standard deviation, as well as the mean reprojection error on the training dataset. Best and second-best results on the test datasets are in green and orange, respectively.

Overall our proposed NS and S-NS camera models showed great performance, yet dependent on the camera and type of enclosure. Notably in the river 4 in dome cases, the S-NS models performed substantially well. These results indicate that the camera placement setup may not be as precise as we believed, but fortunately the camera models were able to compensate for this fault. On the other hand, for river 3 in dome cases, the state-of-the-art KB model outperformed the

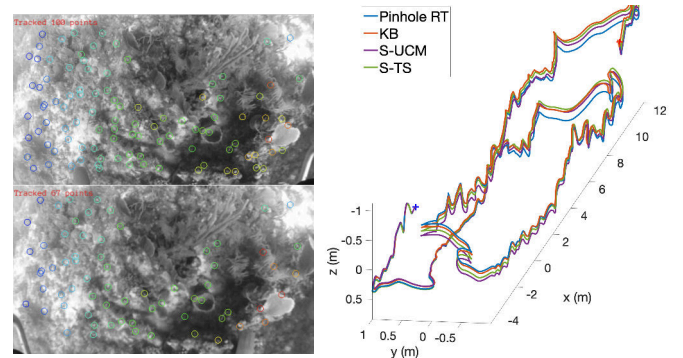


Fig. 3: An example VO [43] run where the CUREE AUV [39] traversed over a reefscape. Here, we used the downward-looking stereo pair of cameras with ArduCam (120°) lenses (left). The plot (right) shows the final trajectories based on the provided camera model.

rest. We observed that the optimized ξ parameters for the NS models were very small, suggesting that the cameras are possibly well-positioned in the dome relative to the z axis and that the NS models, in this case, are only adequate for modeling the fisheye lens distortion. Furthermore, the TS and QS results barely differ from the DS results. While the S-NS models exhibited better results than the NS models, their ξ parameters still did not provide any benefit in representing the underwater image distortions.

We would like to address a general point regarding the number of spheres utilized in the NS camera models. In many cases, adding another sphere to the model, e.g., from TS to QS, does not inherently lead to better performance. For example with the 4 in dome FOCtek (190°) sequence, the QS camera model provided lower training-based mean reprojection error compared to the TS camera model. However, QS testing mean reprojection error was higher than that of TS, signifying that after a certain increase of spheres, the model will lead to overfitting behavior.

B. Visual Odometry Application

We demonstrated the implications of our proposed and state-of-the-art camera models on a *real-time* VO run with the CUREE AUV [39] traversing over a reefscape. In this VO experiment, we first calibrated the downward-looking stereo camera (ArduCam (120°)) intrinsic and extrinsic parameters through *basalt*, and then ran *basalt*’s VO module [43]. Fig. 3 shows the plot of the resulting trajectories based on the following camera models: Pinhole RT, KB, S-UCM (ours), and S-TS (ours). Trajectories based on our proposed models followed closely with the trajectory based on KB. We observed slight offsetting, possibly due to the shifting nature of our models. While it would be interesting to study in more detail with the aid of ground truth data, motion capture systems are not easily deployable in the field. We do consider utilizing controlled experiments in the future.

V. CONCLUSION AND FUTURE WORK

We presented the N-Sphere and Shifted N-Sphere camera models, designed for underwater cameras enclosed in hemispheric dome-ports, a common setup for cameras with wide-angle FOV lenses. The prevalent concerns with this

²<https://cvg.cit.tum.de/research/vslam/basalt>

Dataset Lens Train/Test Ratio #	Test #	KB [28] 8 param.	UCM [35] 5 param.	DS [27] 6 param.	TS * 7 param.	QS * 8 param.	S-UCM * 7 param.	S-DS * 8 param.	S-TS * 9 param.
Air - No dome ADL (185°) 3552 / 1824	1	0.181 (0.217) 0.135	0.387 (0.373) 0.315	0.202 (0.235) 0.147	0.202 (0.233) 0.121	0.185 (0.219) 0.135	0.369 (0.365) 0.301	0.187 (0.224) 0.139	0.187 (0.222) 0.139
	2	0.182 (0.122) 0.138	0.369 (0.266) 0.330	0.172 (0.114) 0.141	0.165 (0.107) 0.140	0.168 (0.110) 0.140	0.356 (0.249) 0.317	0.185 (0.124) 0.142	0.187 (0.127) 0.142
	3	0.175 (0.212) 0.139	0.351 (0.341) 0.325	0.188 (0.222) 0.152	0.188 (0.221) 0.152	0.175 (0.211) 0.139	0.340 (0.329) 0.312	0.177 (0.212) 0.141	0.179 (0.214) 0.141
Air - 4 in dome ADL (185°) 2832 / 1440	1	0.197 (0.177) 0.165	0.255 (0.213) 0.218	0.187 (0.174) 0.153	0.178 (0.167) 0.166	0.196 (0.178) 0.166	0.259 (0.220) 0.218	0.187 (0.173) 0.168	0.177 (0.166) 0.166
	2	0.225 (0.251) 0.149	0.288 (0.283) 0.197	0.215 (0.248) 0.153	0.203 (0.241) 0.151	0.220 (0.251) 0.150	0.296 (0.292) 0.197	0.211 (0.246) 0.153	0.202 (0.241) 0.151
	3	0.193 (0.219) 0.163	0.249 (0.255) 0.221	0.189 (0.217) 0.165	0.181 (0.210) 0.163	0.200 (0.225) 0.162	0.250 (0.255) 0.221	0.194 (0.217) 0.164	0.182 (0.212) 0.161
Pool - 4 in dome ADL (185°) 3504 / 1776	1	0.689 (0.559) 0.594	0.687 (0.539) 0.612	0.701 (0.555) 0.596	0.711 (0.564) 0.596	0.701 (0.548) 0.596	0.685 (0.527) 0.608	0.695 (0.553) 0.593	0.670 (0.549) 0.587
	2	0.694 (0.544) 0.591	0.703 (0.540) 0.604	0.705 (0.549) 0.593	0.711 (0.554) 0.593	0.670 (0.547) 0.592	0.708 (0.537) 0.597	0.698 (0.535) 0.587	0.673 (0.530) 0.579
	3	0.711 (0.548) 0.579	0.725 (0.557) 0.588	0.719 (0.553) 0.580	0.720 (0.554) 0.580	0.703 (0.595) 0.579	0.721 (0.553) 0.585	0.715 (0.551) 0.577	0.699 (0.584) 0.569
River - 4 in dome ADL (185°) 41424 / 21312	1	0.303 (0.271) 0.269	0.327 (0.290) 0.296	0.315 (0.278) 0.272	0.301 (0.270) 0.271	0.302 (0.271) 0.271	0.296 (0.267) 0.260	0.269 (0.246) 0.229	0.266 (0.245) 0.228
	2	0.310 (0.252) 0.264	0.337 (0.275) 0.290	0.325 (0.260) 0.266	0.324 (0.260) 0.266	0.325 (0.260) 0.266	0.299 (0.246) 0.257	0.263 (0.220) 0.231	0.247 (0.211) 0.227
	3	0.311 (0.280) 0.265	0.337 (0.302) 0.290	0.324 (0.287) 0.268	0.323 (0.287) 0.268	0.310 (0.279) 0.266	0.301 (0.277) 0.254	0.273 (0.255) 0.225	0.270 (0.253) 0.223
River - 3 in dome ADL (185°) 45360 / 23328	1	0.219 (0.201) 0.184	0.340 (0.290) 0.290	0.226 (0.205) 0.186	0.221 (0.202) 0.184	0.227 (0.205) 0.186	0.339 (0.290) 0.289	0.224 (0.204) 0.185	0.210 (0.194) 0.179
	2	0.217 (0.197) 0.185	0.343 (0.287) 0.291	0.225 (0.201) 0.187	0.225 (0.201) 0.187	0.226 (0.202) 0.187	0.342 (0.286) 0.290	0.222 (0.199) 0.186	0.224 (0.200) 0.186
	3	0.226 (0.230) 0.179	0.346 (0.314) 0.291	0.233 (0.234) 0.182	0.233 (0.234) 0.182	0.235 (0.234) 0.182	0.344 (0.311) 0.290	0.232 (0.232) 0.180	0.233 (0.234) 0.180
River - 4 in dome FOCtek (190°) 59280 / 30480	1	0.439 (0.534) 0.410	0.440 (0.592) 0.410	0.438 (0.592) 0.410	0.438 (0.592) 0.410	0.442 (0.594) 0.409	0.438 (0.592) 0.408	0.439 (0.593) 0.408	0.439 (0.594) 0.408
	2	0.430 (0.554) 0.416	0.430 (0.553) 0.416	0.429 (0.553) 0.416	0.431 (0.553) 0.416	0.431 (0.553) 0.416	0.427 (0.552) 0.414	0.429 (0.554) 0.414	0.429 (0.554) 0.414
	3	0.438 (0.586) 0.411	0.439 (0.586) 0.411	0.441 (0.586) 0.411	0.437 (0.585) 0.411	0.442 (0.587) 0.411	0.436 (0.585) 0.409	0.438 (0.585) 0.409	0.437 (0.586) 0.409
River - 3 in dome FOCtek (190°) 18000 / 9264	1	0.334 (0.404) 0.308	0.369 (0.446) 0.326	0.370 (0.446) 0.326	0.370 (0.446) 0.326	0.369 (0.445) 0.326	0.351 (0.433) 0.315	0.352 (0.433) 0.315	0.352 (0.434) 0.315
	2	0.330 (0.380) 0.308	0.359 (0.405) 0.334	0.359 (0.406) 0.334	0.359 (0.405) 0.334	0.360 (0.405) 0.334	0.343 (0.393) 0.322	0.343 (0.395) 0.322	0.342 (0.394) 0.322
	3	0.349 (0.421) 0.300	0.379 (0.440) 0.328	0.380 (0.440) 0.328	0.378 (0.439) 0.328	0.379 (0.442) 0.328	0.362 (0.432) 0.315	0.361 (0.431) 0.315	0.361 (0.433) 0.315
Dataset Lens Train/Test Ratio #		Pinhole RT 12 param.	KB [28] 8 param.	UCM [35] 5 param.	DS [27] 6 param.	TS * 7 param.	S-UCM * 7 param.	S-DS * 8 param.	S-TS * 9 param.
Sea - 2 in dome ArduCam (120°) 1144 / 104		0.581 (0.373) 0.479	0.581 (0.368) 0.482	0.574 (0.365) 0.482	0.583 (0.368) 0.481	0.597 (0.378) 0.481	0.571 (0.363) 0.480	0.583 (0.372) 0.480	0.559 (0.357) 0.480

TABLE I: Camera model evaluation across various in-air and in-water scenarios with fisheye (ADL (185°) [36] and FOCtek (190°) [37]) and wide-FOV (ArduCam (120°)) lenses. A * notates our proposed models. We split each dataset uniformly, such that the training dataset consisted of 67% of the data and the remainder assigned to the test dataset. Below the dataset name, we report the number of points used for training and testing, respectively. We report, over ten runs, the mean reprojection error and standard deviation (in pixels) of the test dataset followed by the mean reprojection error (in pixels) of the train dataset. Best and second-best results on the test datasets are in green and orange, respectively.

setup are the refraction effects that occur at the enclosure’s interface and the off-centering of the camera in relation to the dome. Our models address the refraction effect by treating the dome as an additional lens (sphere) in a fisheye-based camera model and compensate for the camera offset by shifting the sphere(s) in the model. We validated our models capability through various calibration experiments in different camera-dome builds and water environments. Furthermore, we showed its effectiveness in a VO application.

An immediate extension of our N-Sphere camera models is to account for the potential rotational offset of the camera with respect to the expected principal axis of a well-positioned camera. This extension would be an important consideration to make, especially if there is any observed camera off-centering. Furthermore, we are interested in conducting a more thorough evaluation of how these various

camera models affect the performance of state-of-the-art VO and VSLAM systems, especially with ORB-SLAM3 [44], which has shown to work well in underwater scenes [17].

A well-represented underwater camera model would have a direct impact on the geometric accuracy for a variety of robotic visual navigation and 3D reconstruction tasks. A vital requirement in operations, such as marine infrastructure inspections, biolife monitoring, and reefscape exploration.

ACKNOWLEDGMENT

We thank S. Jamieson and other members of the WHOI WARPLab for calibration sequence provision. This work is supported by the Link Fellowship, Dartmouth Burke Research Initiation Award, NSF 1919647, 2024541, 2133029.

REFERENCES

- [1] P. Favali, L. Beranzoli, G. D'Anna, F. Gasparoni, J. Marvaldi, G. Clauss, H. W. Gerber, M. Nicot, M. P. Marani, F. Gamberi, et al., "A fleet of multiparameter observatories for geophysical and environmental monitoring at seafloor," *Annals of geophysics*, 2006. 1
- [2] D. Cavallaro and M. Coltelli, "The graham volcanic field offshore southwestern sicily (italy) revealed by high-resolution seafloor mapping and rov images," *Frontiers in Earth Science*, p. 311, 2019. 1
- [3] Y. Liu, M. Hajj, and Y. Bao, "Review of robot-based damage assessment for offshore wind turbines," *Renewable and Sustainable Energy Reviews*, vol. 158, p. 112187, 2022. 1
- [4] N. Sakagami, Y. Yumoto, T. Takebayashi, and S. Kawamura, "Development of dam inspection robot with negative pressure effect plate," *Journal of Field Robotics*, vol. 36, no. 8, pp. 1422–1435, 2019. 1
- [5] N. Gracias, P. Ridaou, R. Garcia, J. Escartín, M. L'Hour, F. Cibecchini, R. Campos, M. Carreras, D. Ribas, N. Palomeras, et al., "The moon: Using a lightweight auv to survey the site of the 17th century ship 'la lune'," in *2013 MTS/IEEE OCEANS-Bergen*. IEEE, 2013, pp. 1–8. 1
- [6] F. Menna, P. Agrafiotis, and A. Georgopoulos, "State of the art and applications in archaeological underwater 3d recording and mapping," *Journal of Cultural Heritage*, vol. 33, pp. 231–248, 2018. 1, 2
- [7] S. B. Williams, O. Pizarro, J. M. Webster, R. J. Beaman, I. Mahon, M. Johnson-Roberson, and T. C. Bridge, "Autonomous underwater vehicle-assisted surveying of drowned reefs on the shelf edge of the great barrier reef, australia," *Journal of Field Robotics*, vol. 27, no. 5, pp. 675–697, 2010. 1
- [8] L. Cai, N. E. McGuire, R. Hanlon, T. A. Mooney, and Y. Girdhar, "Semi-supervised visual tracking of marine animals using autonomous underwater vehicles," *International Journal of Computer Vision*, vol. 131, no. 6, pp. 1406–1427, 2023. 1
- [9] F. S. Hover, R. M. Eustice, A. Kim, B. Englot, H. Johannsson, M. Kaess, and J. J. Leonard, "Advanced perception, navigation and planning for autonomous in-water ship hull inspection," *Int. J. Robot. Res.*, vol. 31, no. 12, pp. 1445–1464, 2012. 1
- [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. 1, 3
- [11] C. Beall, B. J. Lawrence, V. Ila, and F. Dellaert, "3d reconstruction of underwater structures," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2010, pp. 4418–4423. 1
- [12] M. Roznere, P. Mordohai, I. Rekleitis, and A. Q. Li, "3-d reconstruction using monocular camera and lights: Multi-view photometric stereo for non-stationary robots," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 1026–1032. 1
- [13] D. Benham, A. Newman, K. Ellis, R. Gill, and J. G. Mangelson, "3d reconstruction of reefs using autonomous surface vessels and an analysis of chain vs 3d rugosity measurement robustness," in *OCEANS 2022, Hampton Roads*. IEEE, 2022, pp. 1–9. 1
- [14] W. Wang, B. Joshi, N. Burgdorfer, K. Batsosc, A. Q. Lid, P. Mordohai, and I. Rekleitis, "Real-time dense 3d mapping of underwater environments," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 5184–5191. 1
- [15] T. Wang, Q. Zhao, and C. Yang, "Visual navigation and docking for a planar type auv docking and charging system," *Ocean Engineering*, vol. 224, p. 108744, 2021. 1
- [16] A. Quattrini Li, A. Coskun, S. M. Doherty, S. Ghasemlou, A. S. Jagtap, M. Modasshir, S. Rahman, A. Singh, M. Xanthidis, J. M. O'Kane, et al., "Experimental comparison of open source vision-based state estimation algorithms," in *2016 International Symposium on Experimental Robotics*. Springer, 2017, pp. 775–786. 1
- [17] B. Joshi, S. Rahman, M. Kalaitzakis, B. Cain, J. Johnson, M. Xanthidis, N. Karapetyan, A. Hernandez, A. Q. Li, N. Vitzilaios, et al., "Experimental comparison of open source visual-inertial-based state estimation algorithms in the underwater domain," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 7227–7233. 1, 6
- [18] 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. 1, 2
- [19] E. Iscar and M. Johnson-Roberson, "Towards distortion based underwater domed viewpoint camera calibration," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 1896–1902. 1, 2, 3
- [20] C. Balletti, C. Beltrame, E. Costa, F. Guerra, and P. Vernier, "Underwater photogrammetry and 3d reconstruction of marble cargos shipwreck," *The International Archives of the photogrammetry, remote sensing and spatial information sciences*, vol. 40, pp. 7–13, 2015. 1
- [21] R. W. Austin and G. Halikas, "The index of refraction of seawater," 1976. 2
- [22] J.-M. Lavest, G. Rives, and J.-T. Lapresté, "Dry camera calibration for underwater applications," *Machine Vision and Applications*, vol. 13, pp. 245–253, 2003. 2
- [23] T. Łuczynski, M. Pfingsthorn, and A. Birk, "The pinax-model for accurate and efficient refraction correction of underwater cameras in flat-pane housings," *Ocean Engineering*, vol. 133, pp. 9–22, 2017. 2
- [24] F. Menna, E. Nocerino, F. Fassi, and F. Remondino, "Geometric and optic characterization of a hemispherical dome port for underwater photogrammetry," *Sensors*, vol. 16, no. 1, p. 48, 2016. 2
- [25] J. Bosch, N. Gracias, P. Ridaou, and D. Ribas, "Omnidirectional underwater camera design and calibration," *Sensors*, vol. 15, no. 3, pp. 6033–6065, 2015. 2
- [26] C. Silpa-Anan, T. Brinsmead, S. Abdallah, and A. Zelinsky, "Preliminary experiments in visual servo control for autonomous underwater vehicle," in *Proc. IROS*, vol. 4. IEEE, 2001, pp. 1824–1829. 2
- [27] V. Usenko, N. Demmel, and D. Cremers, "The double sphere camera model," in *2018 International Conference on 3D Vision (3DV)*. IEEE, 2018, pp. 552–560. 2, 3, 5, 6
- [28] J. Kannala and S. S. Brandt, "A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 8, pp. 1335–1340, 2006. 2, 3, 5, 6
- [29] J.-M. Lavest, G. Rives, and J.-T. Lapresté, "Underwater camera calibration," in *Computer Vision—ECCV 2000: 6th European Conference on Computer Vision Dublin, Ireland, June 26–July 1, 2000 Proceedings, Part II 6*. Springer, 2000, pp. 654–668. 2
- [30] 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. 2
- [31] P. Corke, C. Detweiler, M. Dunbabin, M. Hamilton, D. Rus, and I. Vasilescu, "Experiments with underwater robot localization and tracking," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*. IEEE, 2007, pp. 4556–4561. 2
- [32] C. Kunz and H. Singh, "Hemispherical refraction and camera calibration in underwater vision," in *OCEANS 2008*. IEEE, 2008, pp. 1–7. 3
- [33] G. Bianco, A. Gallo, F. Bruno, and M. Muzzupappa, "A comparative analysis between active and passive techniques for underwater 3d reconstruction of close-range objects," *Sensors*, vol. 13, no. 8, pp. 11 007–11 031, 2013. 3
- [34] M. Pedersen, S. Hein Bengtson, R. Gade, N. Madsen, and T. B. Moeslund, "Camera calibration for underwater 3d reconstruction based on ray tracing using snell's law," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 1410–1417. 3
- [35] C. Mei and P. Rives, "Single view point omnidirectional camera calibration from planar grids," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*. IEEE, 2007, pp. 3945–3950. 3, 5, 6
- [36] "Adl hfe1414c 1/2" 1.4mm (fisheye) f1.4 fixed iris & manual focus c-mount lens, 5 megapixel rated, 185° fov," <https://www.rmaelectronics.com/adl-hfe1414c/>, Accessed 09/15/2023 2023. 4, 6
- [37] "Foctek cs-m2.5ir(8mp) 1/2.9" (6.4mm image circle) 2.5mm fisheye type f1.6 manual iris & focus cs-mount lens, ir corrected, 8 megapixel rated," <https://www.rmaelectronics.com/foctek-cs-m2-5ir-8mp/>, Accessed 09/15/2023 2023. 4, 6
- [38] "Flir blackfly@s usb3 camera," <https://www.edmundoptics.com/p/bfs-u3-16s2c-cs-usb3-blackflyreg-s-color-camera/40164/>, Accessed 09/15/2023 2023. 4
- [39] Y. Girdhar, N. McGuire, L. Cai, S. Jamieson, S. McCammon, B. Claus, J. E. San Soucie, J. E. Todd, and T. A. Mooney, "Curee: A curious underwater robot for ecosystem exploration," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 11 411–11 417. 5
- [40] "Oak-ffc ov9282 m12," <https://shop.luxonis.com/products/oak-ffc-ov9282-22-pin>, Accessed 09/15/2023 2023. 5
- [41] "Ardumac m23325h12," <https://www.uctronics.com/120-degree-wid>

[e-angle-1-2-3inch-m12-lens-with-lens-adapter-for-raspberry-pi-high-quality-camera.html](#), Accessed 09/15/2023 2023. 5

- [42] T.-T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," *Pattern recognition*, vol. 48, no. 9, pp. 2839–2846, 2015. 5
- [43] V. Usenko, N. Demmel, D. Schubert, J. Stueckler, and D. Cremers, "Visual-inertial mapping with non-linear factor recovery," *IEEE Robotics and Automation Letters (RA-L) & Int. Conference on Intelligent Robotics and Automation (ICRA)*, vol. 5, no. 2, pp. 422–429, 2020. 5
- [44] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "Orb-slam3: An accurate open-source library for visual, visual-inertial, and multimap slam," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021. 6