2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
September 27 - October 1, 2021. Prague, Czech Republic

2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) | 978-1-6654-1714-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/IROS51168.2021.9636198

An Open-Source, Fiducial-Based, Underwater Stereo Visual-Inertial Localization Method with Refraction Correction

Pengfei Zhang, Zhengxing Wu, Jian Wang, Shihan Kong, Min Tan, and Junzhi Yu*, *Fellow, IEEE*

Abstract— **Underwater visual localization is an essential technique for the autonomous operation of underwater robots. However, the unique underwater image characteristics, including refraction, sparse features, and severe noise, pose an enormous challenge to it. For addressing these issues, this paper proposes an open-source fiducial-based underwater stereo visual-inertial localization method under the extended Kalman filter (EKF) framework, which is called FBUS-EKF. First, the refraction is corrected by the refractive camera model and akin triangulation. Second, the fiducial marker and a novel marker pose estimation method are applied to alleviate the adverse effect of sparse features. Third, the EKF is utilized to fuse the inertial and visual information so as to reject the serious noise. Finally, extensive experiments on a test bench demonstrate the effectiveness of the FBUS-EKF method, where the typical localization error is less than 3%, namely, the average error is lower than 3 cm within one meter. The obtained results reveal that the FBUS-EKF method has the prospect to be applied in the precise short-range operation and the localization for underwater robots, which offers a valuable insight for further autonomous underwater task.**

I. Introduction

Nowadays, underwater robots, including bionic robotic fish and autonomous underwater vehicles (AUV), have become an active research area owing to their promising application prospect in marine development. In order to empower them with higher autonomous ability, underwater localization should be tackled due to its essentiality. Compared with the acoustic sensors, like the sonar or Doppler velocity log (DVL), the low-cost visual sensor that can acquire abundant information is more appropriate for the precise short-range operation.

Unlike the high-quality image in the air, the underwater image is usually distorted, feature-sparsed, and quite noisy, resulting in a considerable challenge to the underwater visual localization. More specifically, for underwater scenes, the light is distorted by the refraction effect when it crosses the glass housing and air into the camera. Besides, the limited visual range leads to less visible features, and the suspended particulates as well as light attenuation further introduce the severe noise.

At present, some typical underwater visual localization algorithms have been developed [1]–[8]. For instance, Shkurti *et al.* applied the multi-state constraint Kalman filter (MSCKF) framework, and combined the information of camera, inertial measurement unit (IMU), and depth sensor to estimate the pose of the amphibious robot “Aqua” [1]. Hover *et al.* proposed a pose-graph simultaneous localization and mapping (SLAM) architecture for ship hull inspection, which fuses the data from DVL, image sonar, and monocular camera [2]. Although there are a series of works that have been done, the majority of them neglect the refractive effect and only possess decimeter-level precision. More importantly, all of them are closed-source. Hence, the underwater visual localization still calls for greater effort.

In this paper, we propose an open-source fiducial-based underwater stereo visual-inertial localization method with the consideration of the refractive camera model, which can be applied for the feedback of precise short-range operation as well as the localization in an artificial pool. The main contributions of this study are twofold: 1) A novel underwater fiducial marker pose estimation algorithm with refraction correction is proposed, which significantly improves the accuracy of the traditional marker pose estimation method in the underwater environment. Besides, the proposed FBUS-EKF method further strengthens the robustness of localization, and achieves the centimeter-level precision on the test bench. 2) To the best of our knowledge, the presented work is the first open-source underwater localization method considering the refractive effect, which is publicly available at https://github.com/CASIA-RoboticFish/FBUS-EKF.

II. Related Work

Underwater refraction correction has been exploited for many years. According to the summary of Shortis, the underwater refraction correction techniques can be legitimately classified into three categories, including absorption method, geometric correction, and perspective center shift (PCS) or virtual projection center (VPC) approach [9]. The absorption method is the most common approach, which assumes that the refractive effect can be absorbed by the distortion component of the calibration parameters [10], [11]. This approach is relatively simple, but it neglects the systematic errors induced by the invalid assumption about single projection center [12]. Geometric correction usually applies Snell’s law to trace the light paths through the refractive interfaces, and

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61725305, 62073196, 61633017, and 62022090) and S&T Program of Hebei (Grant No. F2020203037).

P. Zhang, Z. Wu, J. Wang, S. Kong, and M. Tan are with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China and University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: zhangpengfei2017@ia.ac.cn; zhengxing.wu@ia.ac.cn; wangjian2016@ia.ac.cn; kongshihan2016@ia.ac.cn; min.tam@ia.ac.cn).

J. Yu is with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the State Key Laboratory for Turbulence and Complex Systems, Department of Advanced Manufacturing and Robotics, BIC-ESAT, College of Engineering, Peking University, Beijing 100871, China (e-mail: junzhi.yu@ia.ac.cn).

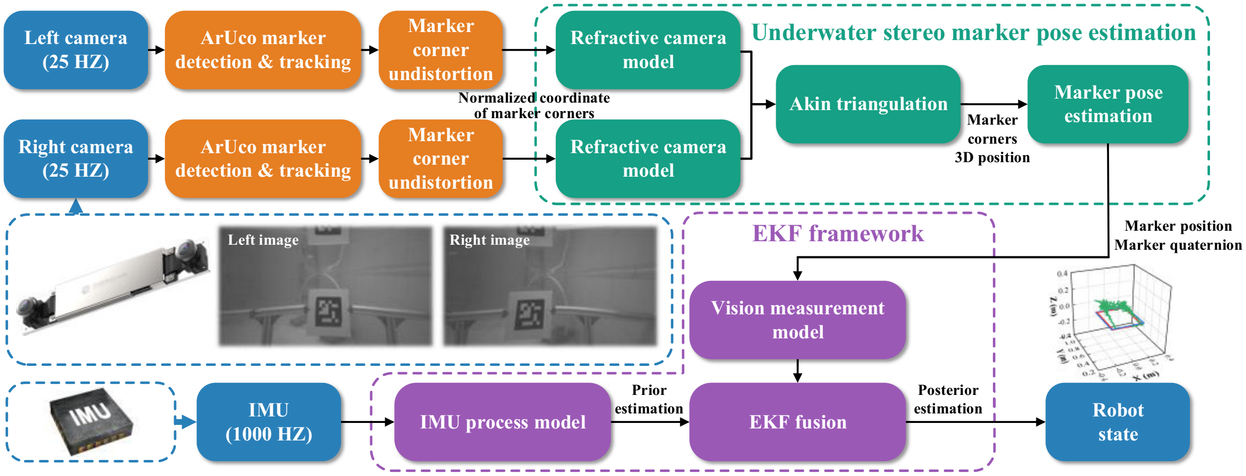


Fig. 1. The schematic of the FBUS-EKF algorithm framework.

finally recovers the real image [13]–[15]. This method is exact, but it is specific to the housing type. PCS or VPC approach is the variation on geometric correction, where the projection model remains unchanged but the projection center is calculated in terms of the refraction principle [16], [17]. This method is effective for both monocular and stereo camera, but it usually needs to introduce extra assumptions. In this paper, the stereo geometric correction approach is applied owing to its high-accuracy and reliability.

A large majority of underwater localization works employ the artificial marker as an aid [18]–[22]. Since the primary use of localization is guiding the operation around the man-made facility, these solutions are also quite viable. Chavez *et al.* employed the monocular camera calibrated by the PinAx refractive model [17] to measure the pose of the fiducial markers, and applied EKF to fuse the data of IMU, DVL, and camera [22]. However, the PinAx model introduces too many assumptions so that it can work effectively only when the camera is very close to the glass housing. In this paper, the ArUco marker is applied for assisting localization due to its robust detecting and tracking performance [23].

To improve the localization performance, visual-inertial fusion is a common technique that compensates for the drift of IMU integration and noisy visual measurement. The mainstream visual-inertial localization method can usually be divided into two categories. A class of methods is based on the Kalman filter, e.g., MSCKF [24], stereo-MSCKF [25], ROVIO [26]. The other spectrum of approaches optimizes the sensor states, formulating the robot localization as a graph optimization problem, e.g., OKVIS [27], VINS-Mono [28]. For better real-time performance, the EKF is utilized as the basic fusion framework here.

III. System Overview

The overview of the proposed FBUS-EKF method is shown in Fig. 1. The overall framework can be divided into three modules. The first part is the processing of visual information. The fiducial markers are detected and tracked from a pair of images based on the open-source library "ArUco", where their correspondence is constructed by means of the marker ID [23]. Then, the marker corners are undistorted applying the fisheye camera model, and the obtained normalized coordinates are inputted into the following operations. The second module recovers the 3D position of corners by the refractive camera model as well as akin triangulation, and calculates the marker pose relative to the camera frame. The third part is the visual-inertial fusion based on EKF. The IMU data are first propagated to calculate a prior estimation of robot state according to the IMU process model, and then the marker pose is applied to update the filter state and acquire posterior estimation through the vision measurement model and EKF fusion.

The notations of this paper are listed here. Five coordinate frames are defined, including world frame $\mathcal{F}_G$, marker frame $\mathcal{F}_M$, IMU or body frame $\mathcal{F}_I$, left camera frame $\mathcal{F}_L$, and right camera frame $\mathcal{F}_R$. The translation vector ${}^{C}\boldsymbol{p}_{AB}$ represents a position vector expressed at $\mathcal{F}_C$, which points from the origin of $\mathcal{F}_A$ to $\mathcal{F}_B$. The quaternion $\boldsymbol{q}_A^B$ and rotation matrix $\boldsymbol{R}_A^B$ denote the rotation of $\mathcal{F}_A$ around $\mathcal{F}_B$. Hence, the vector rotation can be expressed as ${}^{C}\boldsymbol{p}_{AB} = \boldsymbol{q}_A^C \otimes {}^{A}\boldsymbol{p}_{AB} \otimes \boldsymbol{q}_A^{C*}$ or ${}^{C}\boldsymbol{p}_{AB} = \boldsymbol{R}_A^C \cdot {}^{A}\boldsymbol{p}_{AB}$, where $\boldsymbol{q}^*$ is the conjugate of $\boldsymbol{q}$.

IV. Underwater Stereo Marker Pose Estimation

In this section, an underwater stereo marker pose estimation method applied to the air-glass-water flat refractive surfaces is proposed.

A. Refractive Camera Model

The schematic of the refractive camera model and akin triangulation is shown in Fig. 2(a). For the general underwater camera, the complete light path of imaging processes comprises three segments according to various mediums. The purpose of the refractive camera model is to trace the light path of every image pixel in terms of Snell's law and the ray coplanar principle [29].

As shown in Fig. 2(a), the unit vector $\boldsymbol{r}_a$ of the air segment can be easily obtained through the traditional pinhole camera

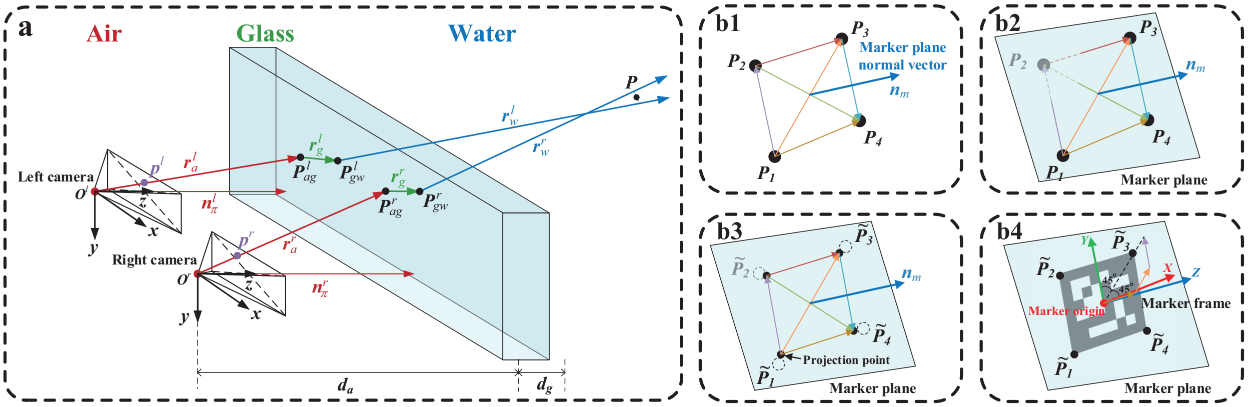


Fig. 2. The schematic of the underwater stereo marker pose estimation algorithm. (a) 3D position reconstruction by refractive camera model and akin triangulation. (b) Calculation procedure of the attitude and position of marker frame.

model [15]. Besides, the intersection point $\boldsymbol{P}_{ag}$ on the air-glass interface can be calculated as

$$\boldsymbol{r}_a = \frac{\overrightarrow{\boldsymbol{Op}}}{||\overrightarrow{\boldsymbol{Op}}||}, \ \boldsymbol{P}_{ag} = \frac{d_a}{\boldsymbol{r}_a \cdot \boldsymbol{n}_\pi}\boldsymbol{r}_a, \tag{1}$$

where $\boldsymbol{p}$ is the normalized coordinates of the object point $\boldsymbol{P}$. $\boldsymbol{n}_\pi$ is the unit normal vector of the glass surface. The vertical distance between the camera optical center and the glass surface is denoted as d_a. Note that the superscript is omitted.

According to the ray coplanar principle, all ray segments from the optical center $\boldsymbol{O}$ to the point $\boldsymbol{P}$ are coplanar with $\boldsymbol{n}_\pi$. Therefore, the glass segment vector $\boldsymbol{r}_g$ can be represented by the linear combination of $\boldsymbol{r}_a$ and $\boldsymbol{n}_\pi$. Similarly, the water segment vector $\boldsymbol{r}_w$ can be obtained as

$$\boldsymbol{r}_g = \alpha_g \boldsymbol{r}_a + \beta_g \boldsymbol{n}_\pi, \tag{2}$$
$$\boldsymbol{r}_w = \alpha_w \boldsymbol{r}_g + \beta_w \boldsymbol{n}_\pi, \tag{3}$$

where

$$\alpha_g = \frac{\mu_a}{\mu_g}, \beta_g = \alpha_g \boldsymbol{r}_a \cdot \boldsymbol{n}_\pi - \sqrt{1 - \alpha_g^2 \left[1 - (\boldsymbol{r}_a \cdot \boldsymbol{n}_\pi)^2\right]},$$
$$\alpha_w = \frac{\mu_g}{\mu_w}, \beta_w = -\alpha_w \boldsymbol{r}_g \cdot \boldsymbol{n}_\pi + \sqrt{1 - \alpha_w^2 \left[1 - (\boldsymbol{r}_g \cdot \boldsymbol{n}_\pi)^2\right]},$$

μ_a, μ_g, and μ_w represent the refractive indexes of air, glass, and water, respectively. Generally, $\mu_a < \mu_w < \mu_g$. At last, the point $\boldsymbol{P}_{gw}$ can be deduced as

$$\boldsymbol{P}_{gw} = \boldsymbol{P}_{ag} + \frac{d_g}{\boldsymbol{r}_g \cdot \boldsymbol{n}_\pi}\boldsymbol{r}_g, \tag{4}$$

where d_g is the thickness of glass.

B. Akin Triangulation

The aim of akin triangulation is to utilize the $\boldsymbol{P}_{gw}$ and $\boldsymbol{r}_w$ from two cameras to recover the 3D position of $\boldsymbol{P}$. First, the vectors from the right camera are represented at $\mathcal{F}_L$ as

$${}^L\boldsymbol{r}_w^r = \boldsymbol{R}_R^L \cdot {}^R\boldsymbol{r}_w^r, \tag{5}$$
$${}^L\boldsymbol{P}_{gw}^r = {}^L\boldsymbol{p}_{LR} + \boldsymbol{R}_R^L \cdot {}^R\boldsymbol{P}_{gw}^r, \tag{6}$$

where the superscript r represents the right camera.

Further, as shown in Fig. 2(a), it can be easily found that the optimal position estimation of object point $\boldsymbol{P}$ is the middle point of the perpendicular line between the ray $\boldsymbol{r}_w^l$ and $\boldsymbol{r}_w^r$. The calculation of middle point position ${}^L\boldsymbol{P}$ can refer to [15].

C. Marker Pose Estimation

In terms of the above methods, the 3D position of the four marker corners can be obtained. In this subsection, the marker position and quaternion with respect to (w. r. t.) $\mathcal{F}_L$ are estimated, which will be applied as the visual measurement of the following filter. Notice that all the vectors and points in this subsection are represented at $\mathcal{F}_L$, and the superscripts are omitted for brevity.

As shown in Fig. 2(b), the overall marker pose estimation comprises four steps. The first step is computing the normal vector of marker plane. Based on the coordinates of four corners, we can define six vectors that point from one corner to the other, e.g., $\boldsymbol{v}_{12} = \overrightarrow{\boldsymbol{P}_1\boldsymbol{P}_2}, \boldsymbol{v}_{13}, \boldsymbol{v}_{14}, \boldsymbol{v}_{23}, \boldsymbol{v}_{24}, \boldsymbol{v}_{34}$. Further, the normal vector can be regarded as the one that is vertical with these six vectors. The optimal normal vector $\boldsymbol{n}_m$ is equivalent to the eigenvector corresponding to the smallest eigenvalue of the matrix $\boldsymbol{M} = \sum \boldsymbol{v}_{ij} \cdot \boldsymbol{v}_{ij}^T$.

The second step is seeking the optimal marker plane equation. Assuming $\boldsymbol{n}_m = [A, B, C]^T$, the marker plane equation can be written as

$$\boldsymbol{\Pi} : Ax + By + Cz + D = 0, \tag{7}$$

where the parameter D is unknown. Generally, the distance between corners and the optimal plane is shortest. Hence, the optimal D_{opt} can be deduced as

$$D_{opt} = -\frac{1}{4}\sum_{i=1}^{4} \boldsymbol{n}_m^T \cdot \boldsymbol{P}_i. \tag{8}$$

The third step is calculating the projection of corners on the marker plane. The projection point $\tilde{\boldsymbol{P}}$ can be deduced as

$$\tilde{\boldsymbol{P}}_i = \boldsymbol{P}_i - t \cdot \boldsymbol{n}_m, \ t = \frac{\boldsymbol{n}_m^T \cdot \boldsymbol{P}_i - \frac{1}{4}\sum_{j=1}^{4} \boldsymbol{n}_m^T \cdot \boldsymbol{P}_j}{||\boldsymbol{n}_m||}. \tag{9}$$

The final step is determining the marker pose relative to $\mathcal{F}_L$. The $\boldsymbol{n}_m$ is defined as the z axis of $\mathcal{F}_M$. Then, the direction of the x axis needs to be determined so that $\mathcal{F}_M$ is fully defined. To reduce the errors, the auxiliary vector $\boldsymbol{m}$ is defined as

$$\boldsymbol{m} = \overline{\boldsymbol{m}}/||\overline{\boldsymbol{m}}||,\ \overline{\boldsymbol{m}} = \overrightarrow{\tilde{\boldsymbol{P}}_1\tilde{\boldsymbol{P}}_2} + \overrightarrow{\tilde{\boldsymbol{P}}_1\tilde{\boldsymbol{P}}_3} + \overrightarrow{\tilde{\boldsymbol{P}}_1\tilde{\boldsymbol{P}}_4}\ . \tag{10}$$

The $\boldsymbol{m}$ is aligned with the angular bisector of the angle between the x axis and y axis, so the x axis can be obtained through rotating $\boldsymbol{m}$ -45 degrees around z axis. According to Rodrigues' rotation formula, the rotation matrix $\overline{\boldsymbol{R}}$ is

$$\overline{\boldsymbol{R}} = \boldsymbol{I} + \sin(-45°)[\boldsymbol{n}_m]_\times + [1 - \cos(-45°)]\,[\boldsymbol{n}_m]_\times^2 \tag{11}$$

Then, the direction vectors of x axis and y axis are

$$\boldsymbol{p}_m = \overline{\boldsymbol{R}} \cdot \boldsymbol{m},\quad \boldsymbol{q}_m = \boldsymbol{n}_m \times \boldsymbol{p}_m. \tag{12}$$

Lastly, the rotation matrix from $\mathcal{F}_M$ to $\mathcal{F}_L$ is obtained as

$$\boldsymbol{R}_M^L = [\boldsymbol{p}_m, \boldsymbol{q}_m, \boldsymbol{n}_m]. \tag{13}$$

The position of the marker frame origin is defined as the average of four corners, which is

$${}^L\boldsymbol{p}_{LM} = (\tilde{\boldsymbol{P}}_1 + \tilde{\boldsymbol{P}}_2 + \tilde{\boldsymbol{P}}_3 + \tilde{\boldsymbol{P}}_4)/4. \tag{14}$$

V. Filter Description

This section describes the filter setup that consists of the process model, measurement model, and EKF fusion [24], [25]. The IMU or robot state is defined as

$$\boldsymbol{X} = \left[{}^G\boldsymbol{p}^T, {}^G\boldsymbol{v}^T, \boldsymbol{q}_I^{G\,T}, \boldsymbol{a}_b^T, \boldsymbol{\omega}_b^T, {}^G\boldsymbol{g}^T\right]^T, \tag{15}$$

where ${}^G\boldsymbol{p} \in \mathbb{R}^3$ and ${}^G\boldsymbol{v} \in \mathbb{R}^3$ represents the position and velocity of $\mathcal{F}_I$ w. r. t. $\mathcal{F}_G$. For brevity, the subscript is omitted. The $\boldsymbol{a}_b \in \mathbb{R}^3$ and $\boldsymbol{\omega}_b \in \mathbb{R}^3$ are the biases of the accelerometer and gyroscope. ${}^G\boldsymbol{g} \in \mathbb{R}^3$ is gravity vector w. r. t. $\mathcal{F}_G$.

A. IMU Process Model

The continuous dynamics for IMU are as follows [30]:

$$\begin{aligned}
{}^G\dot{\boldsymbol{p}} &= {}^G\boldsymbol{v}, \\
{}^G\dot{\boldsymbol{v}} &= \boldsymbol{R}_I^G\left({}^I\boldsymbol{a}_m - \boldsymbol{a}_b - \boldsymbol{a}_n\right) + {}^G\boldsymbol{g}, \\
\dot{\boldsymbol{q}}_I^G &= \frac{1}{2}\boldsymbol{q}_I^G \otimes \left({}^I\boldsymbol{\omega}_m - \boldsymbol{\omega}_b - \boldsymbol{\omega}_n\right), \\
\dot{\boldsymbol{a}}_b &= \boldsymbol{a}_w,\ \dot{\boldsymbol{\omega}}_b = \boldsymbol{\omega}_w,\ {}^G\dot{\boldsymbol{g}} = \boldsymbol{0},
\end{aligned} \tag{16}$$

where the ${}^I\boldsymbol{a}_m$ and ${}^I\boldsymbol{\omega}_m$ are the measurements of accelerometer and gyroscope represented at $\mathcal{F}_I$, respectively. $\boldsymbol{a}_w$, $\boldsymbol{a}_n$, $\boldsymbol{\omega}_n$, and $\boldsymbol{\omega}_w$ are white Gaussian noises.

Besides, the linearized discrete model for the estimated IMU state and error state, as well as the uncertainty propagation are given in [30]. For better accuracy, a 4^{th} order Runge-Kutta numerical intergration is applied to propagate the estimated IMU state.

B. Vision Measurement Model

According to the absolute marker pose and the visual measurements of $\boldsymbol{R}_M^L$ and ${}^L\boldsymbol{p}_{LM}$, the visual measurement equation can be obtained as

$$\begin{aligned}
\boldsymbol{y} &= \boldsymbol{h} + \boldsymbol{v} \\
&= \begin{bmatrix} \boldsymbol{R}_I^L\boldsymbol{R}_G^I\left[{}^G\boldsymbol{p}_{GM_i} - {}^G\boldsymbol{p} - \boldsymbol{R}_I^{G\,I}\boldsymbol{p}_{IL}\right] \\ \boldsymbol{q}_I^L \otimes \boldsymbol{q}_I^{G^*} \otimes \boldsymbol{q}_{M_i}^G \end{bmatrix} + \boldsymbol{v},
\end{aligned} \tag{17}$$

where $\boldsymbol{y} = \left[{}^L\boldsymbol{p}_{LM_i}^T, \boldsymbol{q}_{M_i}^{L\ \ T}\right]^T$ is measurement. M_i represents the marker frame whose ID is i. $\boldsymbol{q}_{M_i}^L$ is the quaternion corresponding to $\boldsymbol{R}_{M_i}^L$. ${}^G\boldsymbol{p}_{GM_i}$ and $\boldsymbol{q}_{M_i}^G$ are the known information. $\boldsymbol{v} \in \mathbb{R}^7$ is the white Gaussian noise with covariance $\boldsymbol{R}$, where $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R})$.

Based on (17), the Jacobian matrix w. r. t. the error state can be defined as follows:

$$\boldsymbol{H} = \begin{bmatrix} -\boldsymbol{R}_I^L\boldsymbol{R}_G^I & \boldsymbol{O} & \boldsymbol{R}_I^L\left[\boldsymbol{R}_G^I({}^G\boldsymbol{p}_{GM_i} - {}^G\boldsymbol{p})\right]_\times & \boldsymbol{O}_{3\times9} \\ \boldsymbol{O} & \boldsymbol{O} & [\boldsymbol{q}_{M_i}^G]_R[\boldsymbol{q}_I^L]_L\boldsymbol{L}_2[\boldsymbol{q}_I^G]_L\boldsymbol{L}_1 & \boldsymbol{O}_{4\times9} \end{bmatrix}, \tag{18}$$

where

$$\boldsymbol{L}_1 = \frac{1}{2}\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \boldsymbol{L}_2 = \frac{\partial \boldsymbol{q}_I^{G^*}}{\partial \boldsymbol{q}_I^G} \approx \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix},$$

$[\boldsymbol{q}]_L$ and $[\boldsymbol{q}]_R$ denote the left- and right- quaternion-product matrices, respectively [30].

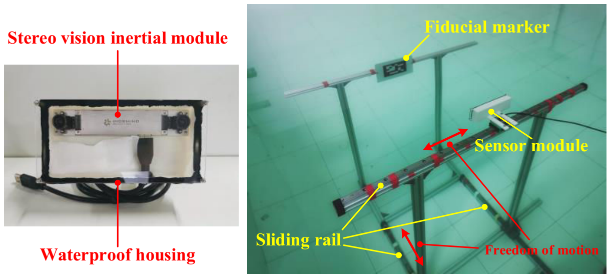
C. EKF Fusion

The EKF fusion process consists of time update and measurement update stages. The ordinary form of EKF can be found in [30]. In our implementation, the time update stage is carried out when the new visual results come, and a batch of IMU data are processed at once. Furthermore, the nearest marker detected from the images is applied to measurement update. Besides, the marker poses w. r. t. the world frame $\mathcal{F}_G$ are known in advance, which are constant and not updated in the iterations.

VI. Experiments

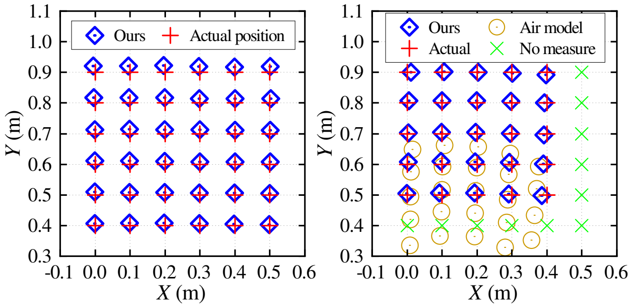
In order to evaluate the proposed method, two experiments were performed in this section. In these experiments, the sensor suite called Indemind stereo vision inertial module was applied, which collected the monochrome image with resolution 640 × 480 pixels at 25 Hz and the IMU data at 1000 Hz (see Fig. 3(a)). Note that the time synchronization of sensor data is implemented in the hardware layer. All algorithms ran on a laptop with an Intel Core i5-7300HQ processor and 8 GB RAM, whose real-time performance is completely satisfied for the online localization. Besides, the adopted marker size was 16 cm.

In addition, measuring the actual camera trajectory is a great challenge in the underwater environment. Hence, for obtaining the precise ground truth, the experiments were conducted on the test bench that consists of the two-directional sliding rails, where the sensor suite can move



(a) (b)

Fig. 3. The sensor module and test bench.



(a) Air case (b) Underwater case

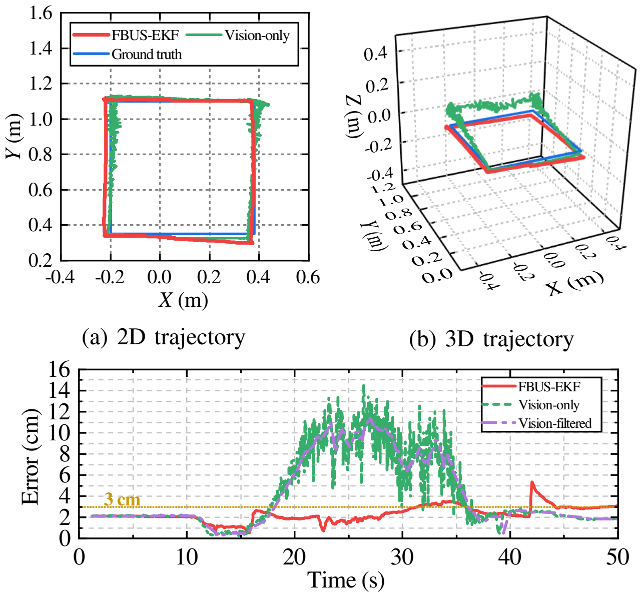
Fig. 4. The results of marker position estimation.

freely in a horizontal plane (see Fig. 3(b)). The geometric parameters of the sliding rails are completely known, so the camera position and trajectory along the sliding rails can be measured accurately.

A. Marker Pose Estimation Experiment

The precise marker pose estimation is the foundation of the following visual-inertial fusion. In this experiment, the sensor module was placed in various locations to collect a period of image data about a fixed marker, then the average marker pose estimation results were calculated by the proposed method. Note that the sensor module was mounted at a slider on the test bench and moved along sliding rails.

Fig. 4 depicts the position estimation results on the air and water. Note that there are some positions that were not measured in the underwater case due to the reduced view caused by the refractive effect. Besides, when the d_a and d_g are set as zero, the proposed underwater method can degenerate into the air case. It can be easily found that the estimated positions of ours method are quite close to actual positions, and the overall Root Mean Square Error (RMSE) of position is lower than 2 cm no matter in the air case or the underwater case. However, when the pinhole camera model on the air is directly employed in the underwater environment, the estimated positions seriously deviate from the actual one. This is because underwater cameras in flat port housing are in fact axial cameras rather than pinhole cameras [29]. Hence, this experiment validates the necessity and accuracy of the proposed method.



(a) 2D trajectory (b) 3D trajectory

(c) Localization error

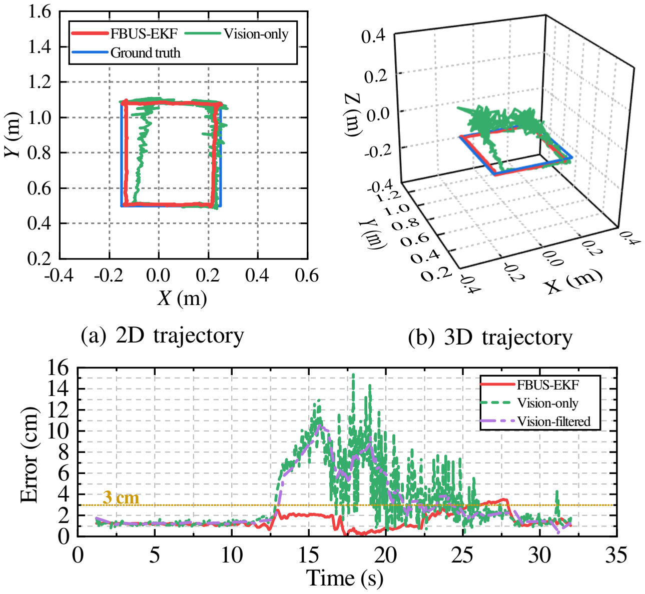
Fig. 5. The localization results in the air.

B. Visual-Inertial Localization Experiment

To overcome the precision degradation caused by the severe measurement noise in water and the increasing distance from marker, the combination of IMU and camera becomes very imperative. In this experiment, the sensor module was moved along a rectangular trajectory on the test bench, and a marker was fixed in front of the sensor for localization. Then, the localization accuracy of three cases was compared, including the proposed FBUS-EKF, the vision-only method, and the vision-filtered method which smooths the results of the vision-only method by a mean filter.

Figs. 5 and 6 show the trajectories and error curves of the air case and underwater case, respectively. The coordinate origin represents the position of the marker. The estimated trajectories of all methods are close to the ground truth when the sensor is near the marker. However, with the increasing of the distance between sensor and marker, especially for the underwater case, the vision-only and vision-filtered method seriously deviate from the actual trajectory. This deviation is caused by the poor attitude estimation accuracy. In general, when the attitude error is 5° and the distance from marker is 120 cm, the localization error will reach about 10 cm, which can be easily calculated by $\sin(5^\circ) \times 120 \approx 10.45$ cm.

Based on the above analysis, the vision-only method is hardly available for practical application when the attitude error can not be restrained effectively. Similarly, the vision-filtered method only reduces the fluctuation of the trajectory but cannot decrease the localization error at the source. Nevertheless, for the FBUS-EKF, the IMU data provides more motion information, and thus the estimated trajectory is pretty consistent with the ground truth, whose average localization error is lower than 3 cm for both air and underwater cases. In a word, this experiment fully demonstrates



(a) 2D trajectory

(b) 3D trajectory

(c) Localization error

Fig. 6. The localization results in the water.

the effectiveness and superiority of the FBUS-EKF compared with the vision-only method.

VII. Conclusions and Future Work

In this paper, we have presented an open-source underwater localization method for the precise short-range operation and the localization in an artificial pool, which achieves the centimeter-level precision on the test bench. Owing to its open-source property and well precision, this work will create convenience for the studies about the autonomous operation and advanced control of underwater robots.

In the future, we will further focus on the localization in the scene that the marker pose is unknown, and the robust marker detection algorithm for the camera with fast shaking.

References

[1] F. Shkurti, I. Rekleitis, M. Scaccia, and G. Dudek, “State estimation of an underwater robot using visual and inertial information,” in *Proc. IEEE Int. Conf. Intell. Rob. Syst.*, San Francisco, United States, Sep. 2011, pp. 5054–5060.

[2] F. S. Hover *et al.*, “Advanced perception, navigation and planning for autonomous in-water ship hull inspection,” *Int. J. Robot. Res.*, vol. 31, no. 12, pp. 1445–1464, 2012.

[3] A. Kim and R. M. Eustice, “Real-time visual SLAM for autonomous underwater hull inspection using visual saliency,” *IEEE Trans. Robot.*, vol. 29, no. 3, pp. 719–733, 2013.

[4] W. Wang and G. Xie, “Online high-precision probabilistic localization of robotic fish using visual and inertial cues,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 2, pp. 1113–1124, 2014.

[5] S. Rahman, A. Q. Li, and I. Rekleitis, “Sonar visual inertial slam of underwater structures,” in *Proc. IEEE Int. Conf. Robot. Autom.*, Brisbane, Australia, May. 2018, pp. 5190–5196.

[6] S. Rahman, A. Q. Li, and I. Rekleitis, “SVIn2: An underwater SLAM system using sonar, visual, inertial, and depth sensor,” in *Proc. IEEE Int. Conf. Intell. Rob. Syst.*, Macau, China, Nov. 2019, pp. 1861–1868.

[7] C. Gu, C. Yang, and G. Sun, “Environment driven underwater camera-IMU calibration for monocular visual-inertial SLAM,” in *Proc. IEEE Int. Conf. Robot. Autom.*, Montreal, Canada, May. 2019, pp. 2405–2411.

[8] Y. Wang, X. Ma, J. Wang, and H. Wang, “Pseudo-3D vision-inertia based underwater self-localization for AUVs,” *IEEE Trans. Veh. Technol.*, vol. 69, no. 7, pp. 7895–7907, 2020.

[9] M. Shortis, “Calibration techniques for accurate measurements by underwater camera systems,” *Sensors*, vol. 15, no. 12, pp. 30810–30827, 2015.

[10] R. Li, C. Tao, W. Zou, R. G. Smith, and T. A. Curran, “An underwater digital photogrammetric system for fishery geomatics,” *Int. Arch. Photogramm. Remote Sens.*, vol. 31, pp. 317–323, 1996.

[11] A. Meline, J. Triboulet, and B. Jouvencel, “A camcorder for 3D underwater reconstruction of archeological objects,” in *Proc. OCEANS*, Seattle, United States, Sep. 2010, pp. 1–9.

[12] A. Sedlazeck and R. Koch, “Perspective and non-perspective camera models in underwater imaging-overview and error analysis,” *Outdoor and Large-Scale Real-World Scene Analysis*, Berlin, Germany: Springer, pp. 212–242, 2012.

[13] R. Li, H. Li, W. Zou, R. G. Smith, and T. A. Curran, “Quantitative photogrammetric analysis of digital underwater video imagery,” *IEEE J. Ocean. Eng.*, vol. 22, no. 2, pp. 364–375, 1997.

[14] A. Jordt-Sedlazeck and R. Koch, “Refractive calibration of underwater cameras,” in *Proc. Eur. Conf. Comput. Vis.*, Florence, Italy, Oct. 2012, pp. 846–859.

[15] S. Kong, X. Fang, X. Chen, Z. Wu, and J. Yu, “A NSGA-II-based calibration algorithm for underwater binocular vision measurement system,” *IEEE Trans. Instrum. Meas.*, vol. 69, pp. 794–803, 2020.

[16] G. Telem and S. Filin, “Photogrammetric modeling of underwater environments,” *ISPRS J. Photogramm. Remote Sens.*, vol. 65, no. 5, pp. 433–444, 2010.

[17] T. Łuczyński, M. Pfingsthorn, and A. Birk, “The pinax-model for accurate and efficient refraction correction of underwater cameras in flat-pane housings,” *Ocean Eng.*, vol. 133, no. 3, pp. 9–22, 2017.

[18] M. Carreras, P. Ridao, R. García, and T. Nicosevici, “Vision-based localization of an underwater robot in a structured environment,” in *Proc. IEEE Int. Conf. Robot. Autom.*, Taipei, China, Sep. 2003, pp. 971–976.

[19] D. Kim, D. Lee, H. Myung, and H. Choi, “Artificial landmark-based underwater localization for AUVs using weighted template matching,” *Intelligent Serv. Robot.*, vol. 7, no. 3, pp. 175–184, 2014.

[20] A. D. Buchan, E. Solowjow, D. Dueckera, and E. Kreuzer, “Low-cost monocular localization with active markers for micro autonomous underwater vehicles,” in *Proc. IEEE Int. Conf. Intell. Rob. Syst.*, Vancouver, Canada, Sep. 2017, pp. 4181–4188.

[21] J. Jung, J. Li, H. Choi, and H. Myung, “Localization of AUVs using visual information of underwater structures and artificial landmarks,” *Intelligent Serv. Robot.*, vol. 10, no. 1, pp. 67–76, 2017.

[22] A. G. Chavez, C. A. Mueller, T. Doernbach, and A. Birk, “Underwater navigation using visual markers in the context of intervention missions,” *Int. J. Adv. Robot. Syst.*, vol. 16, no. 2, pp. 1–14, 2019.

[23] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. Marín-Jiménez, “Automatic generation and detection of highly reliable fiducial markers under occlusion,” *Pattern Recognit.*, vol. 47, no. 6, pp. 2280–2292, 2014.

[24] A. I. Mourikis and S. I. Roumeliotis, “A multi-state constraint Kalman filter for vision-aided inertial navigation,” in *Proc. IEEE Int. Conf. Robot. Autom.*, Roma, Italy, Apr. 2007, pp. 3565–3572.

[25] K. Sun *et al.*, “Robust stereo visual inertial odometry for fast autonomous flight,” *IEEE Robot. Autom. Lett.*, vol. 3, no. 2, pp. 965–972, 2018.

[26] M. Bloesch, M. Burri, S. Omari, M. Hutter, and R. Siegwart, “Iterated extended Kalman filter based visual-inertial odometry using direct photometric feedback,” *Int. J. Robot. Res.*, vol. 36, no. 10, pp. 1053–1072, 2017.

[27] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, “Keyframe-based visual–inertial odometry using nonlinear optimization,” *Int. J. Robot. Res.*, vol. 34, no. 3, pp. 314–334, 2015.

[28] T. Qin, P. Li, and S. Shen, “VINS-Mono: A robust and versatile monocular visual-inertial state estimator,” *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004–1020, 2018.

[29] A. Agrawal, S. Ramalingam, Y. Taguchi, and V. Chari, “A theory of multi-layer flat refractive geometry,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, United States, Jun. 2012, pp. 3346–3353.

[30] J. Sola, “Quaternion kinematics for the error-state KF,” *Laboratoire d-Analyse et dArchitecture des Systemes-Centre national de la recherche scientifique (LAAS-CNRS) Toulouse, France, Tech. Rep.*, 2012.