

Article

# Study of Underwater Large-Target Localization Based on Binocular Camera and Laser Rangefinder

Wenbo Xu <sup>1,2,3,4</sup>, Xinhui Zheng <sup>1,2,3,4</sup>, Qiyao Tian <sup>1,2,4,\*</sup> and Qifeng Zhang <sup>1,2,4,5</sup>

- <sup>1</sup> State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China; xubowen@sia.cn (W.X.); zhengxinhui@sia.cn (X.Z.); zqf@sia.cn (Q.Z.)  
<sup>2</sup> Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110169, China  
<sup>3</sup> Key Laboratory of Marine Robotics, Shenyang 110169, China  
<sup>4</sup> University of Chinese Academy of Sciences, Beijing 100049, China  
<sup>5</sup> Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou 511458, China  
\* Correspondence: tianqiyao@sia.cn

**Abstract:** Currently, for underwater close-range large-target localization, visual localization techniques fail since large targets completely occupy the camera's field of view at ultraclose ranges. To address the issue, a multi-stage optical localization method combining a binocular camera and a single-point laser rangefinder is proposed in this paper. The proposed method comprises three parts. First, the imaging model of the underwater camera is modified, and a laser rangefinder is used to further correct the underwater calibration results of the binocular camera. Second, YOLOv8 is applied to recognize the targets to prepare for target localization. Third, extrinsic calibration of the binocular camera and laser rangefinder is performed, and a Kalman filter is employed to fuse the target position information measured by the binocular camera and laser rangefinder. The experimental results show that, compared with using a binocular camera alone, the proposed method can accurately and stably locate the target at close ranges with an average error of only 2.27 cm, without the risk of localization failure, and reduces binocular localization error by 90.57%.

**Keywords:** underwater large-target localization; camera calibration; laser rangefinder; target recognition



**Citation:** Xu, W.; Zheng, X.; Tian, Q.; Zhang, Q. Study of Underwater Large-Target Localization Based on Binocular Camera and Laser Rangefinder. *J. Mar. Sci. Eng.* **2024**, *12*, 734. <https://doi.org/10.3390/jmse12050734>

Academic Editor: Mohamed Benbouzid

Received: 23 March 2024  
Revised: 21 April 2024  
Accepted: 24 April 2024  
Published: 28 April 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

With the continuous exploration of marine environments, underwater object detection technology has been widely used in autonomous navigation [1], obstacle avoidance of underwater vehicles [2], and various underwater detection tasks [3].

Most underwater localization technologies rely on acoustic signals, optical signals, and electromagnetic signals. In acoustic localization [4,5], electroacoustic conversion is used to measure the propagation time and orientation of acoustic waves based on the propagation characteristics of acoustic waves in water. These measurements are used to determine the position of underwater objects. Sonar has a long detection range and is not affected by water turbidity [6]. In electromagnetic localization [7,8], underwater objects are localized by measuring the abnormal magnetic field generated by the superposition of underwater magnetic material in the geomagnetic field. This method has a unique advantage in detecting underwater low-noise submarines, iron, or strong magnetic objects buried on the sea floor. Optical localization mainly refers to visual localization [9]. In this approach, optical imaging is applied to obtain target images, and a visual locating algorithm is used to obtain the position of targets in the images, which are high resolution and contain rich information. Visual localization has become an important technology for underwater object observation, recognition, and manipulation.

However, the popular technologies cannot fully satisfy the requirements for locating large underwater targets at close ranges, probably within 5 m, and they all have their own

limitations. Acoustic localization is commonly used for medium- and long-range target localization, and its detection range is correlated with the frequency of the sound wave and ranges from hundreds to thousands of meters. Electromagnetic localization can be interrupted by an existing environmental magnetic field, and this technique is suitable for the rough localization of targets [6]. Optical localization can achieve high-precision localization at close ranges underwater. However, large targets at ultraclose ranges will fill the camera's field of view, resulting in visual localization failure. Inspired by autonomous driving [10] on land, we incorporate a single-point laser rangefinder to address the issue that visual localization fails at ultraclose ranges. A camera is used to recognize and locate large targets at close ranges, and a laser rangefinder is used to locate large targets at ultraclose ranges.

Many studies have been conducted on underwater visual localization. Zhang et al. [11] estimated the relative distance between a spherical target and a monocular camera by extracting the radius dimension of the target contour in the image plane. Xu et al. [12] proposed a method of relative position estimation using multiple ArUco markers [13], in which the information extracted by multiple markers was fused to improve the localization accuracy. Arturo et al. [14] identified augmented reality (AR) markers to obtain the relative positions of underwater vehicles and markers. Li et al. [15] proposed a redundant localization method that combined monocular and binocular vision. They first used the P3P (perspective three-point) question [16] to locate the array light source via monocular vision, then used binocular vision to locate the light source again, and finally fused the information regarding the two positions. Wang et al. [17] used two cameras to locate an underwater circular target according to the binocular ranging principle. Meng et al. [18] extracted color features to recognize a target and then located the target through binocular vision. Zhong et al. [19] applied binocular vision to identify three centrosymmetric navigation lamps and estimate relative positions.

Few studies have been conducted on underwater laser applications. Chris et al. [20] used a CCD camera and two-line lasers for distance measurement of multiple positions of underwater targets in a scenario where acoustic and visual detection were difficult. Takashi [21] et al. used a camera and two single-point lasers to measure the distance between the target and the ocean floor and estimate the pitching state of an underwater vehicle. In the literature [22–24], the application of an underwater laser scanner was described; however, this scanner greatly differed from a laser rangefinder.

The visual localization method using navigation lamps and markers is mainly suitable for underwater autonomous docking, in which the lamps or markers are placed in advance. In this study, the position of the target is unknown, and no markers or lamps are used. A binocular camera can achieve target localization without external objects. Therefore, a binocular camera and a single-point laser rangefinder are selected as the optical localization devices for underwater vehicles for large, close-range targets. The main innovations of this study can be summarized as follows:

- A method that utilizes a laser rangefinder to aid in calibrating the camera is designed to significantly enhance the accuracy of binocular camera calibration.
- A multi-stage optical localization method that combines a binocular camera and a laser rangefinder is designed to achieve stable and accurate target localization.

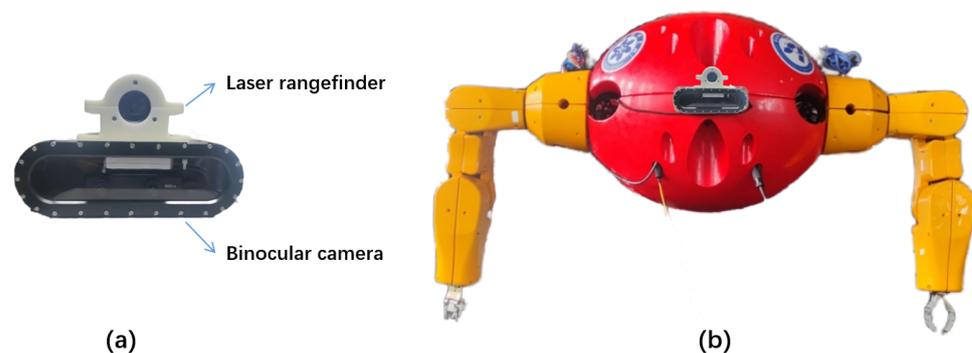
The remainder of this study is organized as follows: Section 2 describes the underwater optical localization system and the different stages of underwater close-range large-target localization. Section 3 details the specific implementation of the optical localization method for close-range large-target localization. Section 4 presents the experimental results of underwater close-range large-target localization. Section 5 concludes the paper.

## 2. Design of the Optical Localization System

### 2.1. Description of the Optical Localization System

The underwater optical localization system is composed of a target perception module and an image processing module. As shown in Figure 1a, the target perception module

contains a binocular camera and a laser rangefinder. The baseline length of the binocular camera is 63 mm, the focal length is 2.8 mm, and the maximum field of view is  $90^\circ(H) \times 60^\circ(V) \times 100^\circ(D)$ . The measuring range of the binocular camera on land is 0.1 m–15 m. Different wavelengths absorb light to different degrees underwater, with the blue-green wavelengths absorbing the least underwater [25]. In the literature [26], the power of a red laser is about ten times that of a green laser, but the high absorption of the red wavelength severely limits the maximum detection range. Considering this effect, underwater lidar also selects blue-green light as the operating wavelength to effectively minimize the absorption of light by water [27]. Therefore, the green band wavelength is selected for the laser rangefinder in the study. The measuring frequency of the laser rangefinder is 3 Hz, the measuring range is 0.05 m–5 m, and the measuring accuracy is  $\pm 3$  mm. The image processing module uses a Jetson Xavier industrial computer to deploy the deep convolutional neural network model YOLOv8. In the optical localization system, the video stream collected by the binocular camera is transmitted to the Jetson Xavier processor through a USB cable, which is used to detect the target, the relative position of the target is calculated, and the position information is published through the ROS node. The laser rangefinder returns distance information and this information is also published through the ROS node. The companion computer receives the target information obtained from the binocular camera and laser rangefinder for further processing.



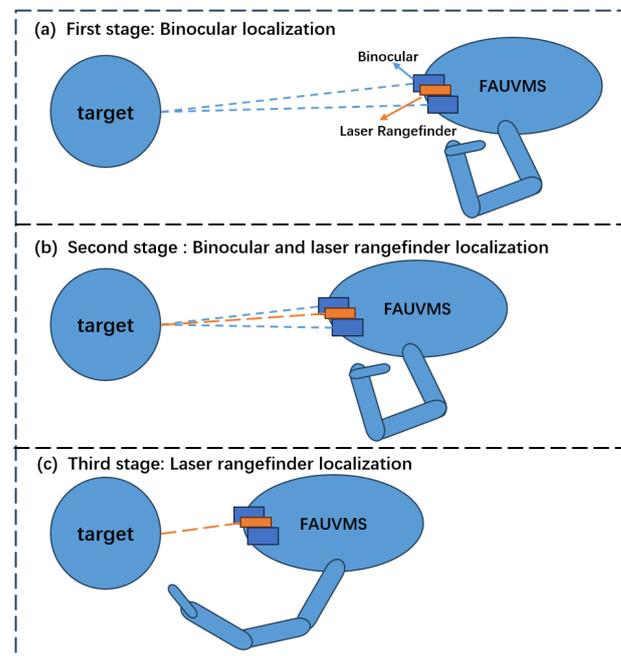
**Figure 1.** (a) Optical localization equipment. (b) Prototype of FAUVMS.

The localization system is mounted on an underwater vehicle manipulator system (FAUVMS). FAUVMS, developed by the Shenyang Institute of Automation, is a novel underwater intervention system with manipulators as its core. It validates underwater interventions from vehicle-based to manipulator-based operations. It consists primarily of an underwater vehicle and two heavy-load electric manipulators with a forward operating range of up to 630 mm, as shown in Figure 1b. The underwater trajectory tracking and manipulation performance has been demonstrated in previous experiments. More details about FAUVMS are provided in [28]. The intervention task for FAUVMS is to recognize (primarily large) underwater targets and manipulate them at close ranges. In this case, distinctive features are lacking because the target fully fills the field of view of the binocular camera in the manipulation process; as a result, the binocular camera cannot locate the target and visual localization fails. A novel underwater localization method is urgently needed to address this issue.

## 2.2. Different Target Localization Stages

The localization process of the optical localization system for a close-range large target is shown in Figure 2. In the first stage, after the target enters the field of view of the binocular camera, FAUVMS begins to recognize and locate the target and adjust its attitude to make the binocular camera face the target. In the second stage, as FAUVMS gradually approaches the target, the laser rangefinder also begins to work, and the binocular camera and laser rangefinder jointly locate the target. In the third stage, when the large target fills the field of view of the binocular camera, the binocular camera fails to locate the target. At

this point, the laser rangefinder independently locates the target until FAUVMS reaches the operating area of the target.



**Figure 2.** Three stages of underwater large-target localization. (a) Binocular camera locates the target alone. (b) Binocular camera and laser rangefinder locate the target together. (c) Laser rangefinder locates the target alone.

### 3. Methods

The specific implementation of the optical localization method proposed in this study is shown in Figure 3. First, considering the influence of the waterproof cover and water, the binocular camera is calibrated underwater. Next, a laser rangefinder is used to correct the underwater calibration results to further improve the underwater binocular ranging accuracy. Then, the image dataset of the target is collected underwater. The deep learning method YOLOv8 is adopted to teach the model to recognize the target and provide pixel coordinates for subsequent target localization. The localization of large underwater close-range targets is divided into three stages. In the first stage, the binocular camera is used to locate the target independently. In the second stage, the binocular camera and laser rangefinder are used to jointly locate the target. To align the coordinate systems of the binocular camera and laser rangefinder, the laser rangefinder and binocular camera are extrinsically calibrated. The target position information obtained by the binocular and laser rangefinder is fused using the Kalman filter (KF). In the third stage, the laser rangefinder locates the target independently. Finally, experiments are conducted in the pool to verify the effectiveness of the scheme.

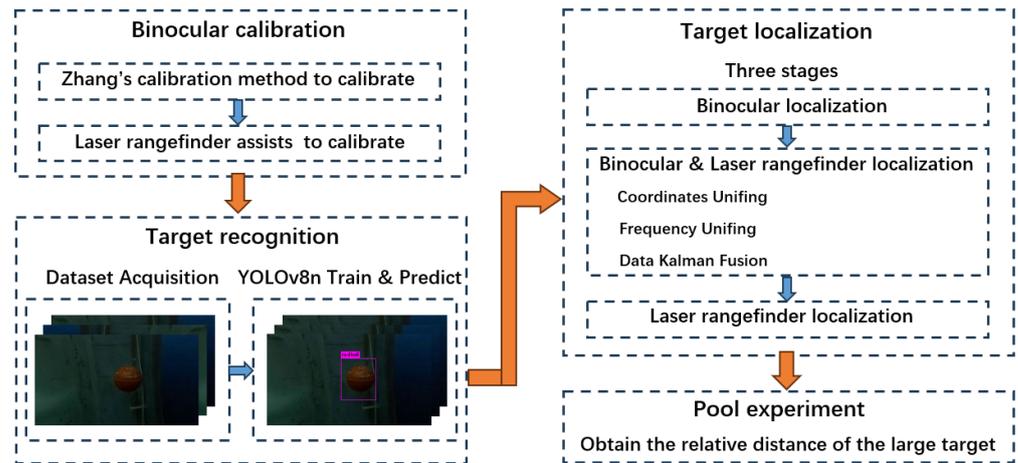


Figure 3. The implementation flow of the proposed method.

### 3.1. Laser Rangefinder-Assisted Camera Calibration

The relationship between the three-dimensional information of any point in space and the corresponding point in the image can be obtained by calibrating the binocular camera. The camera calibration accuracy directly determines the reliability of the binocular ranging results. However, the underwater binocular camera often requires a waterproof layer, which causes light to refract many times during propagation, so the pinhole imaging model used on land is not suitable for underwater imaging. In this study, the underwater imaging model is modified to conform to the pinhole imaging model, and then a laser rangefinder is applied to further correct the camera parameters.

#### 3.1.1. Camera Imaging Model

The pinhole camera model is commonly adopted to represent the imaging principle of cameras. The imaging models of monocular and binocular cameras are the same. As shown in Figure 4a, the size of the image changes as the pinhole plane moves forward and backward. The pinhole camera model is shown as follows:

$$x = f \frac{X}{Z} \tag{1}$$

where  $f$  represents the focal length of the camera,  $Z$  represents the distance from the hole to the object,  $X$  represents the length of the object, and  $x$  represents the length of the object on the image plane.

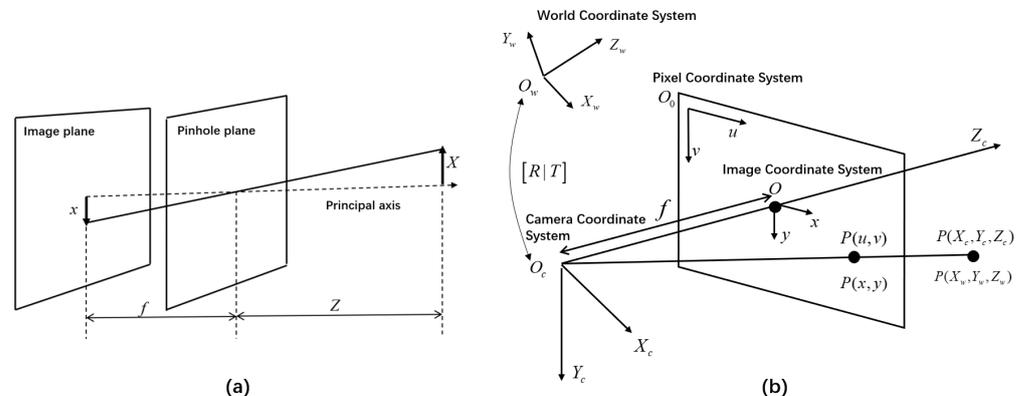


Figure 4. (a) The pinhole camera model. (b) The relationship diagram of the coordinate systems in the camera.

For convenience, the transformation process from the world coordinate system to the pixel coordinate system is represented in Figure 4b. Transformation between the world coordinate system and the camera coordinate system can be achieved through coordinate system rotation and translation, as shown in Equation (2).

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{2}$$

where  $R$  represents the  $3 \times 3$  rotation matrix,  $T$  represents the translation vector, and  $P(X_c, Y_c, Z_c)$  and  $P(X_w, Y_w, Z_w)$  represent world coordinates and camera coordinates of point  $P$ , respectively. The relationship between the camera coordinates and image coordinates is shown in Equation (3), which can be expressed in matrix form, as shown in Equation (4).

$$\begin{aligned} x &= \frac{f_x X_c}{Z_c} \\ y &= \frac{f_y X_c}{Z_c} \end{aligned} \tag{3}$$

$$Z_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & 0 & 0 \\ 0 & f_y & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \tag{4}$$

where  $P(x, y)$  represent image coordinates of point  $P$ , and  $(f_x, f_y)$  represent the focal length of the camera. The relationship between the image coordinates and pixel coordinates is shown in Equation (5), which can be expressed in matrix form, as shown in Equation (6).

$$\begin{aligned} u &= \frac{x + c_x \times d_x}{d_x} \\ v &= \frac{y + c_y \times d_y}{d_y} \end{aligned} \tag{5}$$

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{d_x} & 0 & c_x \\ 0 & \frac{1}{d_y} & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{6}$$

where  $P(u, v)$  represent pixel coordinates of point  $P$ , and  $(c_x, c_y)$  represent the pixel coordinates in the center of the image. Finally, the transformation relationship between the world coordinate system and the pixel coordinate system can be derived as follows:

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{f_x}{d_x} & 0 & c_x & 0 \\ 0 & \frac{f_y}{d_y} & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = M_1 M_2 \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{7}$$

where  $M_1$  and  $M_2$  represent the internal and external parameter matrices of the camera, respectively.

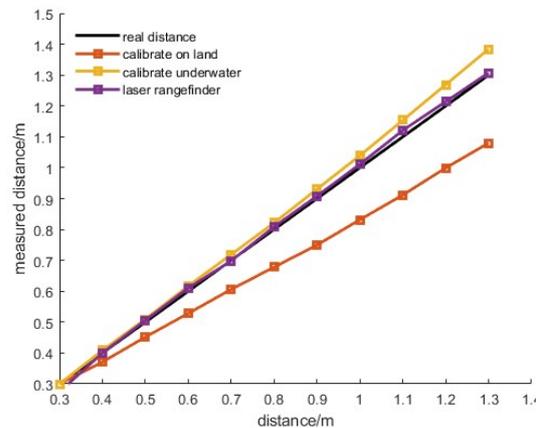
### 3.1.2. Modified Model for Underwater Imaging

Zhang’s calibration method [29] is a popular, highly accurate camera calibration method that is often used in land camera calibration. This method is used to calibrate the binocular camera in this study. The appropriate acquisition number of the checkerboard is 18 pairs [30]. To improve the calibration results, 64 pairs of images are collected on land and underwater. The calibration results for the binocular camera on land and underwater are shown in Table 1.

**Table 1.** Camera intrinsic parameters.

Parameters	$f_x/d_x$	$f_y/d_y$	$c_x$	$c_y$	$k_1$	$k_2$
$Left_{air}$	1401.8	1400.9	936.03	541.67	−0.1656	0.0012
$Right_{air}$	1403.6	1402.97	940.06	530.88	−0.1647	−0.0077
$Left_{water}$	1869.3	1868.2	941.22	540.66	0.0852	−0.1497
$Right_{water}$	1866.94	1866.36	938.94	531.8	0.084	−0.1158

In Table 1, only  $f_x/d_x$  and  $f_y/d_y$  change greatly between the calibration results on land and those underwater except for  $k_1$  and  $k_2$ , which are radial distortion parameters used to eliminate image distortion.  $d_x$  and  $d_y$  are the physical size of a single pixel, which does not change, so the parameter that changes the most is the focal length. To compare the effects of the land and underwater calibration results on binocular ranging, the land and underwater camera calibration results are used for underwater ranging. The binocular ranging is performed within 0.1 m to 1.3 m. Each distance is measured five times and the standard variance approximates 0. The ranging results are shown in Figure 5.



**Figure 5.** Ranging results under different conditions. Black represents the true value, orange represents the underwater ranging result using the land calibration result, yellow represents the underwater ranging result using the underwater calibration result, and purple represents the underwater ranging result using the laser rangefinder.

In Figure 5, the calibration results on land are applied to underwater binocular ranging, and the ranging error is very large and grows linearly. The reason is the binocular camera is not affected by light refraction when calibrated on land, but underwater light refraction cannot be ignored. When the underwater calibration results are applied to underwater binocular ranging, the ranging error is clearly lower, but it still increases as the distance increases and also grows linearly. When the distance is 1.3 m, the ranging error reaches 0.1 m.

Ideally, the slopes of these three lines should match the slope of the black line, but in practice, only the slope of the laser rangefinder always approaches the ideal slope. A comparison between the calibration results for land and underwater water reveals that the most important parameter is the focal length. Moreover, in the literature [31], the largest difference between the camera calibration results after considering the multilayer refraction of underwater light and the calibration results obtained by using Zhang’s calibration method is also the focal length. Therefore, the focal length is the key factor for improving binocular underwater ranging.

To correct the focal length of the camera, the underwater imaging model of the camera is analyzed. The main difference between binocular camera on land and underwater is that binocular cameras must be placed in waterproof covers when used underwater. Light

enters the waterproof cover from the water and then enters the camera from the waterproof cover, undergoing two refractions. Therefore, the underwater imaging model no longer conforms to the pinhole camera model. To ensure that the underwater imaging model remains consistent with the pinhole camera model, the underwater imaging model is modified as shown in Figure 6. The waterproof cover is only 2 mm thick, and the refraction of the waterproof cover to the light can be ignored because the thin waterproof cover does not change the light's final direction and only induces a negligible radial shift [32].

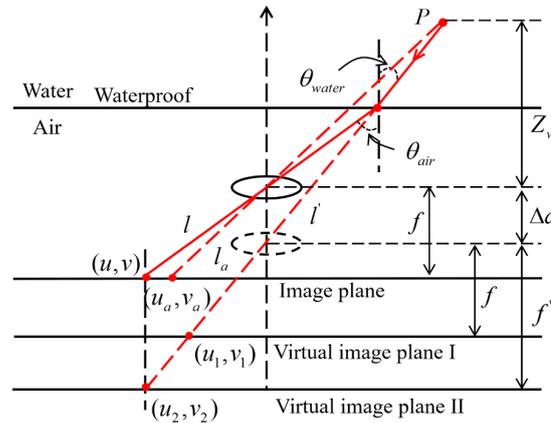


Figure 6. Underwater imaging model.

In Figure 6, the light starts from the three-dimensional point  $P$  and is affected by water refraction. The pixel coordinates of point  $P$  on the real image plane are  $(u, v)$ . If refraction is not considered, the pixel coordinates of point  $P$  on virtual image plane I are  $(u_1, v_1)$ . The pixel coordinates obtained without considering refraction are clearly inconsistent with the actual pixel coordinates. Therefore, the dotted line  $l_1$  is extended to find the coordinates  $(u_2, v_2)$  that agree with the coordinates  $(u, v)$ . The current focal length changes and is called the virtual focal length  $f'$ . Then, the sources of binocular ranging errors are further analyzed. The schematic diagram of the binocular ranging principle is shown as in Figure 7.

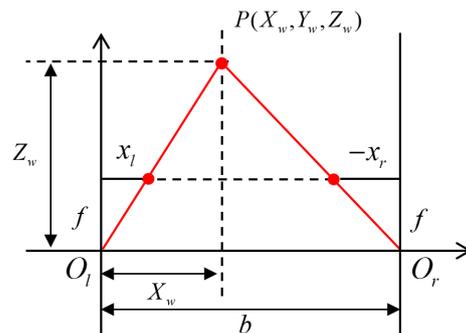


Figure 7. The schematic diagram of binocular ranging principle.

In Figure 7, the optical centers of the left and right cameras are  $O_l$  and  $O_r$ . The distance between  $O_l$  and  $O_r$  is the baseline, denoted as  $b$ .  $x_l$  and  $-x_r$  are the  $x$  values of point  $P$  in the left and right image coordinate systems, respectively.

According to the triangle similarity principle, we can obtain

$$\frac{x_l}{f} = \frac{X_w}{Z_w} \tag{8}$$

$$\frac{x_r}{f} = \frac{b - X_w}{Z_w} \tag{9}$$

Then, we further obtain

$$Z_w = \frac{f \times b}{x_l - x_r} \tag{10}$$

where  $(x_l - x_r)$  represents the difference between the  $x$  values of point  $P$  in the image coordinate system of the left and right cameras, called the disparity.

Equation (10) shows that  $Z_w$  in world coordinates is directly related to disparity and focal length. The camera calibration is calculated according to the pinhole camera mode and does not take water refraction into account. In Figure 6, point  $P$  corresponds to  $(u_a, v_a)$  according to this calibration; however, it actually corresponds to  $(u, v)$ . The incorrect pixel coordinates cause the incorrect image coordinates. The disparity calculated from the image coordinates directly affects the calculation of  $Z_w$ . Therefore, the binocular camera still has error after underwater calibration. However, in Figure 6, considering the effect of refraction to approximate the underwater imaging model as a pinhole model, a virtual focal length  $f'$  is constructed so that the  $P$  point correctly corresponds to  $(u, v)$ , which can eliminate the error.

Therefore, the calculation of the unknown  $f'$  is key. The modified underwater imaging model is shown as

$$\begin{aligned} \frac{Z_w + \Delta d}{f'_x} &\approx \frac{Z_w}{f'_x} = \frac{X_w}{x_u} \\ \frac{Z_w + \Delta d}{f'_y} &\approx \frac{Z_w}{f'_y} = \frac{Y_w}{y_v} \end{aligned} \tag{11}$$

According to Equation (5),  $f'_x$  and  $f'_y$  can be further expressed as

$$\begin{aligned} f'_x &= 1000 \times \frac{Z_w}{X_w} (u - c_x) d_x \\ f'_y &= 1000 \times \frac{Z_w}{Y_w} (v - c_y) d_y \end{aligned} \tag{12}$$

where  $(X_w, Y_w, Z_w)$  represent the world coordinates of point  $P$ ,  $P(x_u, y_v)$  represent the image coordinates corresponding to the pixel coordinates  $P(u, v)$ ,  $\Delta d$  represents the distance difference between the actual lens and the virtual lens, which is very small and can be ignored, and  $f'_x$  and  $f'_y$  represent the virtual focal lengths in the  $x$  and  $y$  directions, respectively.

$P(X_w, Y_w, Z_w)$  is difficult to obtain using a binocular camera. However, it can be obtained more easily by using a laser rangefinder and combining the position relationship between the laser rangefinder and the binocular camera. The detailed derivation of the position relationship is described in Section 3.3.1. Finally, the virtual focal length can be obtained to correct the camera calibration result.

### 3.2. Underwater Target Recognition

Target recognition is the key step in target localization. In the traditional target recognition method, the target is recognized by learning from artificially extracted features. This method exhibits strong real-time performance but can be easily impacted by different environments. In target recognition based on deep learning, information from pixel-level raw data to abstract semantic concepts is extracted layer by layer, which gives this method an outstanding advantage in extracting global features and context information from images. As a result, this method is greatly adaptable to target recognition in different environments.

The YOLO [33–35] series is used to input the image directly into the detection model and output the results. The main advantages of this series are its simple structure, small size, and fast speed. The most recent YOLO version is YOLOv8 [36], which has a higher recognition accuracy, smaller size, and faster speed than the previous version. Due to its limited computing resource requirements and high real-time performance, lightweight YOLOv8n is selected for the target recognition method in this study.

The dataset for underwater targets is collected in different pools and under different light conditions, as shown in Figure 8a. The state of the target in the dataset is relatively rich. The aspect ratio, position, and size of the target pixel in the dataset are constantly changing. The dataset is randomly divided into a training set, a test set, and a validation set at a ratio of 8:1:1. The number of training rounds is set to 50, and the number of single training images is set to 4. The dataset is enhanced using the mosaic method.

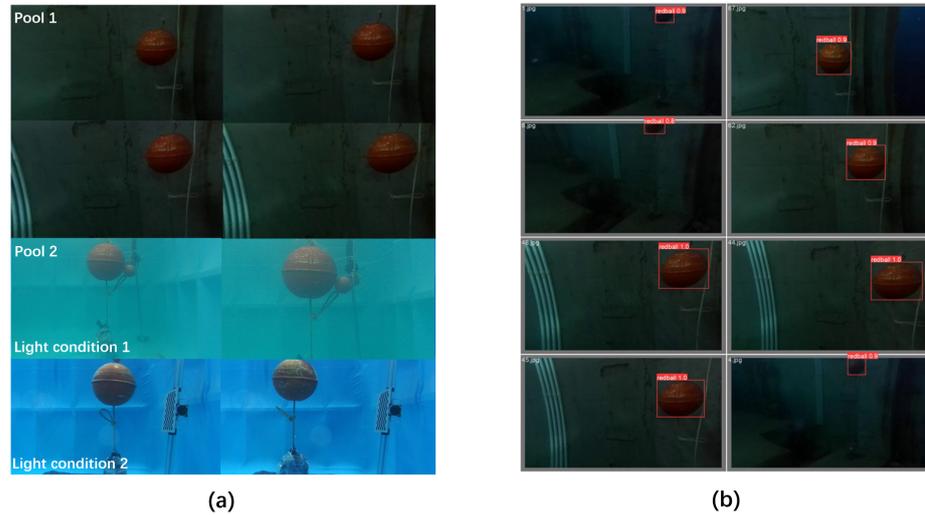


Figure 8. (a) Dataset for underwater target. (b) The prediction results of the test set.

The prediction results of the test set are shown in Figure 8b. The target can still be recognized under poor underwater conditions. The target precision–recall curve plot is shown in Figure 9. Precision and recall are calculated as follows

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$
(13)

where  $TP$  represents the number of samples that are predicted to be positive and are actually positive,  $FP$  represents the number of samples that are predicted to be positive but are actually negative, and  $FN$  represents the number of samples that are predicted to be negative but are actually positive.

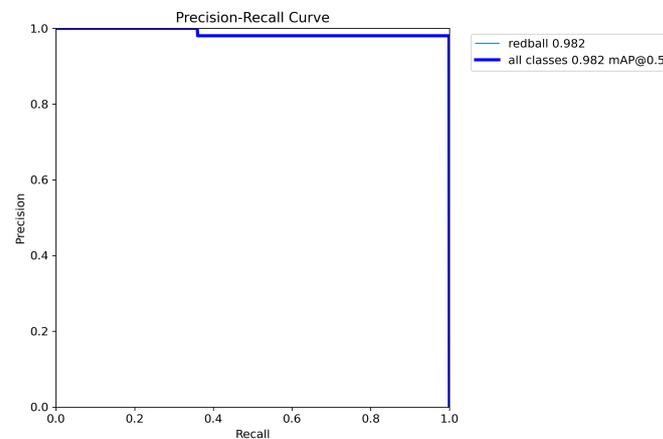
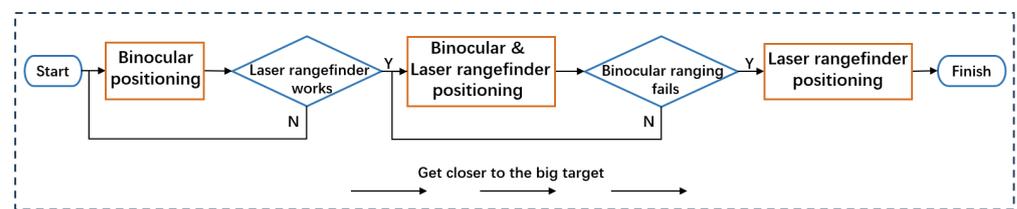


Figure 9. Precision versus recall curve. Since there is only one class of target, the P-R curve of the target coincides with the P-R curves of all classes.

In this study, *AP* is used to evaluate the recognition effect of the YOLOv8 model. *AP* is the average accuracy of different recall rates and is represented on the PR curve as the area enclosed by the PR curve. The *IoU* is the intersection of the real box and the predicted box. *IoU* = 0.5 indicates that the target is considered detected at *IoU* greater than 0.5. As shown in Figure 9, when *IoU* is 0.5, the *AP* of YOLOv8 for target recognition reaches 0.982, which shows that YOLOv8 performs well in underwater target recognition.

### 3.3. Localization of Large Targets Underwater

The localization of close-range large targets by the optical localization system consists of three stages. The flow chart of this process is shown in Figure 10. The condition of entering the second stage is determined by whether the laser rangefinder works, and the condition of entering the third stage is determined by whether the binocular ranging fails. The localization schemes of the three stages can be switched smoothly to effectively complete close-range large-target localization by the optical localization system.



**Figure 10.** The localization process of large targets underwater.

#### 3.3.1. Extrinsic Calibration of Binocular Camera and Laser Rangefinder

To achieve multisensor fusion, the coordinate systems of the binocular camera and laser rangefinder must be consistent. However, the relative position relationship between the binocular camera and laser rangefinder is unknown, and coordinate transformation cannot be performed directly. After the binocular camera and laser rangefinder are fixed, whether on land or underwater, their position relationship does not change. The binocular camera is not accurate underwater but is accurate on land. Therefore, we determine the transformation relationships of the binocular camera and laser rangefinder on land. The laser rangefinder leaves a laser spot on the target. This spot can be detected by the binocular camera as a feature point, and the world coordinates of the point can thereby be obtained. The position of the laser spot is represented in the coordinate system of the laser rangefinder and the binocular camera, so the transformation relationship between the two coordinates can be identified.

The laser spot recognition method is shown in Figure 11. Due to the high brightness of the laser spot, the image is transferred from the BGR color space to the gray space and flipped, and the gray value is used to recognize the laser spot for the first time. The laser spot is approximately a circle, so circularity is used for the second recognition. The pixel area of the laser spot is within a fixed range, so the pixel area size of the laser spot is used for the third recognition. As shown in Figure 12, the laser spot can be stably recognized by the binocular camera underwater by using this method.

According to the relationship between the monocular camera and laser rangefinder in the literature [37], we establish a relationship diagram between the binocular camera and laser rangefinder, as shown in Figure 13a. The laser spot on the checkerboard can be recognized by the binocular camera; thus, the camera coordinates of the laser spot can be obtained. The laser rangefinder returns only one distance value, *L*. As shown in Figure 13b, the distance value of the laser rangefinder can be transformed into coordinates in the binocular camera coordinate system. The camera coordinates of the laser spot are expressed as

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} L\cos\theta_x + t_x \\ L\cos\theta_y + t_y \\ L\cos\theta_z + t_z \end{bmatrix} \tag{14}$$

where  $(X_c, Y_c, Z_c)$  represent the camera coordinates of the laser spot,  $(\cos\theta_x, \cos\theta_y, \cos\theta_z)$  represent the angles between the laser ray and the axis of the camera coordinate, and  $(t_x, t_y, t_z)$  represent the translation vector between the camera origin and the laser rangefinder. As shown in Figure 13b,  $\cos\theta_x, \cos\theta_y,$  and  $\cos\theta_z$  satisfy Equation (15).

$$\cos^2\theta_x + \cos^2\theta_y + \cos^2\theta_z = 1 \tag{15}$$

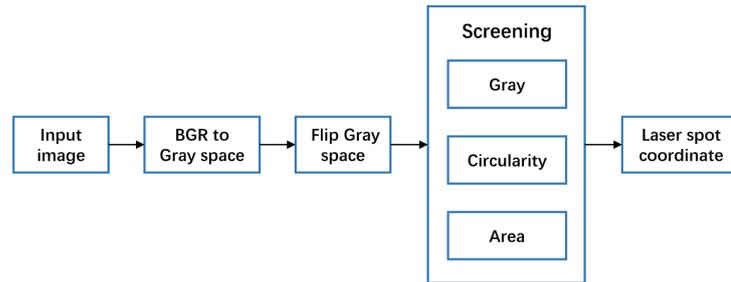


Figure 11. Recognition process of laser spot.

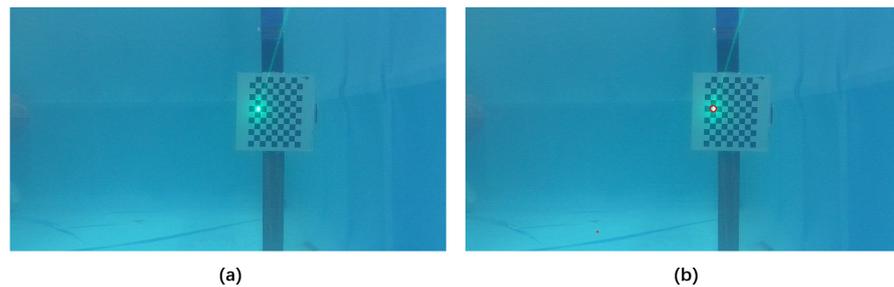


Figure 12. (a) Raw laser spot underwater. (b) Laser spot recognized underwater.

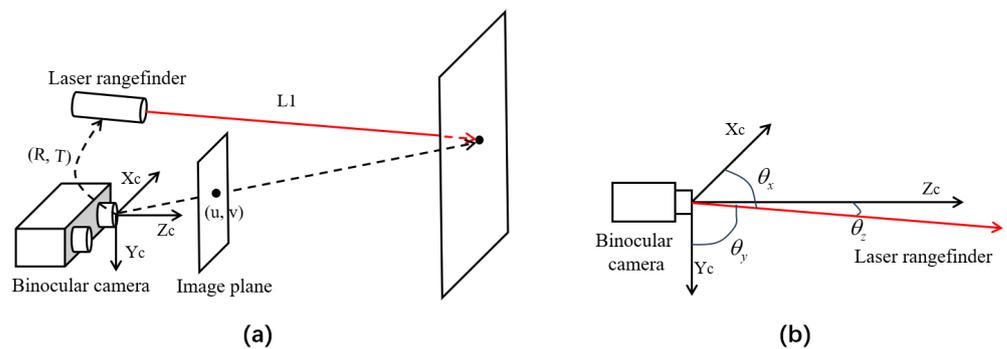


Figure 13. (a) Relation diagram of the binocular camera and laser rangefinder. (b) Relation diagram of the laser rangefinder after translation transformation and binocular camera.

Therefore, the camera coordinates of the laser spot are further expressed as

$$\begin{bmatrix} L\cos\theta_x + t_x \\ L\cos\theta_y + t_y \\ L\sqrt{1 - \cos^2\theta_y - \cos^2\theta_z} + t_z \end{bmatrix} \tag{16}$$

In Figure 13a, the pixel coordinates of the laser spot are  $(u, v)$ . According to Equations (4) and (6), the relationship between the distance  $L$  and the pixel coordinates  $(u, v)$  is expressed as

$$\frac{f_x(L\cos\theta_x + t_x)}{(L\sqrt{1 - \cos^2\theta_y - \cos^2\theta_z + t_z})d_x} + c_x = u \tag{17}$$

$$\frac{f_y(L\cos\theta_y + t_y)}{(L\sqrt{1 - \cos^2\theta_y - \cos^2\theta_z + t_z})d_y} + c_y = v \tag{18}$$

where  $(f_x, f_y)$  represents the focal length of the camera,  $(d_x, d_y)$  represent the physical size of a single pixel in the  $x$  and  $y$  directions, and  $(c_x, c_y)$  represent the pixel coordinates of the center of the image. At different distances where the laser spot can be recognized by the camera, Equations (17) and (18) are stacked, and the unknowns  $(\cos\theta_x, \cos\theta_y, \cos\theta_z, t_x, t_y, t_z)$  are solved.

### 3.3.2. Distance Estimation Based on the Binocular Camera and Laser Rangefinder

- First stage of target localization.

The principle of binocular ranging is shown in Figure 7. According to Figure 7 and to Equation (10), we continue to further obtain

$$X_w = \frac{Z_w \times x_l}{f}; Y_w = \frac{Z_w \times y_l}{f} \tag{19}$$

where  $f$  represents the focal length of the camera, and  $y_l$  is the  $y$  values of point  $P$  in the image coordinate systems of the left camera, and  $P(X_w, Y_w, Z_w)$  represent the world coordinates of point  $P$ .

In this study, the left camera coordinate system is set as the world coordinate system, and  $P(X_w, Y_w, Z_w)$  and  $P(X_c, Y_c, Z_c)$  are the same. Therefore, the distance estimation of the first stage can be expressed as

$$D_1 = \sqrt{X_w^2 + Y_w^2 + Z_w^2} \tag{20}$$

The center of the target anchor frame in target recognition is taken as the target positioning point in this stage. To prevent a single-point measurement from deviating, the point and its four neighbors are measured simultaneously. The median distance of these five points is considered the distance.

- Second stage of target localization.

In this stage, where the binocular camera and laser rangefinder work together, FAU-VMS acquires two types of distance information about the target. The measurement principle of the laser rangefinder is based on the time difference between laser emission and reception, which differs from the principle of the binocular camera. Therefore, to locate the target more reliably, the distances obtained by the binocular camera and laser rangefinder are fused by the KF. Before distance fusion, the localization frequency of the binocular camera must be kept consistent with that of the laser rangefinder. The distances measured by the binocular camera and laser rangefinder are set to  $D_{21}$  and  $D_{22}$ , respectively, and  $k$  is the Kalman gain in the range  $[0, 1]$ . Then, the fused distance  $D_2$  is shown as

$$D_2 = D_{21} + k(D_{22} - D_{21}) \tag{21}$$

The standard deviation of  $D_2$  is set to  $\sigma$ , and the standard deviations of  $D_{21}$  and  $D_{22}$  are set to  $\sigma_1$  and  $\sigma_2$ , respectively. Then, the variance of  $D_2$  can be expressed as

$$\sigma^2 = Var(D_{21} + k(D_{22} - D_{21})) \tag{22}$$

Considering that  $D_{21}$  and  $D_{22}$  are independent of each other, we further obtain

$$\sigma^2 = (1 - k)^2 \times \sigma_1^2 + k^2 \times \sigma_2^2 \tag{23}$$

To minimize the variance of  $D_2$ , the derivative of Equation (23) with respect to  $k$  is set to 0. The Kalman gain  $k$  is solved as

$$k = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \tag{24}$$

Finally, the fused distance  $D_2$  of the binocular camera and laser rangefinder is calculated as

$$D_2 = D_{21} + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} (D_{22} - D_{21}) \tag{25}$$

Notably, the binocular camera takes the laser spot on the target as the target positioning point in this stage. The distance returned by the laser rangefinder can be fused with the distance calculated by the binocular camera only after this distance is converted into camera coordinates.

- Third stage of target localization.

In this stage, the laser rangefinder is used to estimate the distance of the target. This process also requires converting the distance returned by the laser rangefinder into camera coordinates and using Equation (20) to calculate the distance to the target, as shown

$$D_3 = \sqrt{(L\cos\theta_x + t_x)^2 + (L\cos\theta_y + t_y)^2 + (L\cos\theta_z + t_z)^2} \tag{26}$$

#### 4. Experiment and Analysis

The laser rangefinder-assisted camera calibration experiment and the large-target localization experiment are conducted in a standing pool that is 3 m × 2 m × 1.2 m with no disturbances. The positions of the binocular camera and laser rangefinder are fixed to each other. The underwater calibration of the binocular camera is completed, and the initial calibration parameters are obtained. The laser attenuates underwater, and different water qualities affect the degree of laser attenuation. Therefore, after the laser rangefinder is placed in water, it must be corrected. To ensure the convenience of relative distance measurement during the experiment, FAUVMS is fixed in the pool.

##### 4.1. Experiments on Camera Calibration Assisted by Laser Rangefinder

The distances returned by the laser rangefinder and the world coordinates of the laser spot acquired by the binocular camera are collected at 0.5 m, 1.0 m, 1.5 m, 2.0 m, and 2.5 m on land.  $L$  and spot coordinates are all measured five times and then the average is taken. Some data are shown in Table 2. The unknowns can be solved using Equations (3) and (4), as shown in Table 3.

**Table 2.** Results of laser rangefinder and binocular camera at different distances on land.

Distance	0.5 m	1.0 m	1.5 m	2.0 m
$L$	0.501 m	1.002 m	1.510 m	2.008 m
Spot coordinates (m)	(−0.026, −0.012, 0.494)	(−0.053, −0.024, 1.006)	(−0.153, 0.032, 1.529)	(−0.196, 0.011, 2.048)

**Table 3.** Transition parameters of laser rangefinder and binocular camera.

Parameters	$\cos \theta_x$	$t_x$	$\cos \theta_y$	$t_y$	$\cos \theta_z$	$t_z$
Result	−0.061	−0.03935	−0.026	0.08826	0.998	−0.19866

In the pool, we measure the distance of the checkerboard at different distances five times. Finally, we correct the laser rangefinder until it works in the pool. The underwater ranging effect of the laser rangefinder is shown in Table 4, and the ranging error is very small.

**Table 4.** Ranging results of laser rangefinder after underwater correction.

Distance	1.2 m	1.4 m	1.6 m	1.8 m	2.0 m	2.2 m
Error	−0.0059 m	−0.0084 m	−0.0050 m	0.0013 m	−0.0012 m	0.0007 m

The distance returned by the laser rangefinder is transformed into coordinates in the world coordinate system underwater according to Equation (14). The binocular camera is used to recognize the laser spot and obtain the pixel coordinates of the spot. According to Equation (12), the focal lengths  $f'_x$  and  $f'_y$  can be calculated from the world and pixel coordinates of the laser spot. These lengths are calculated every 0.1 m within 1.0 m to 2.2 m. The calculated results are partially shown in Table 5.

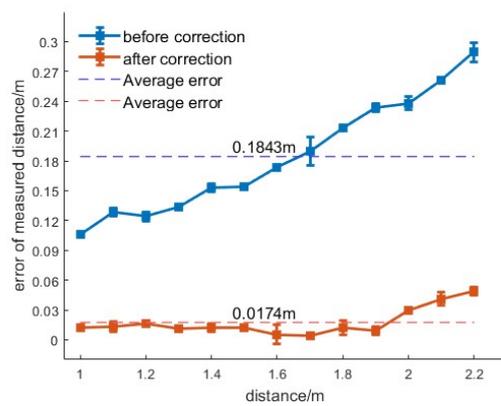
**Table 5.** Focal lengths calculated at different distances.

Distance	1.2 m	1.4 m	1.6 m
World coordinates (m)	(−0.1117, 0.0569, 1.2071)	(−0.1239, 0.0516, 1.4098)	(−0.1357, 0.0465, 1.6066)
Pixel coordinates	(544, 231)	(541, 244)	(542, 246)
$f'_x$ (mm)	3.1956	3.2282	3.4064
$f'_y$ (mm)	3.3061	2.8386	3.3136
Distance	1.8 m	2.0 m	2.2 m
World coordinates (m)	(−0.1473, 0.0415, 1.8005)	(−0.1594, 0.0362, 2.0032)	(−0.1713, 0.0311, 2.2015)
Pixel coordinates	(537, 251)	(533, 255)	(536, 261)
$f'_x$ (mm)	3.2726	3.1638	3.3895
$f'_y$ (mm)	3.2940	3.3169	2.5458

In Table 5,  $f'_x$  and  $f'_y$  are roughly distributed around a certain value. The recognized range of the laser spot is a circular area. The world coordinates of the laser spot are obtained using the center of this area as the pixel coordinates. However, the pixel coordinates of the laser spot are not always accurate; the changed world coordinates significantly affect  $f'_x$  and  $f'_y$ , which results in a scattered distribution of  $f'_x$  and  $f'_y$ . Therefore, the focal lengths in Table 5 are averaged to obtain the final corrected focal lengths  $f'_x = 3.25$  mm and  $f'_y = 3.26$  mm. Underwater binocular ranging is performed within 1.0 m to 2.2 m underwater using the corrected focal lengths. Each distance is measured eight times and the ranging error is calculated. The results are shown in Table 6.  $Error_b$  and  $Error_a$  represent the binocular ranging errors before and after correction, respectively. The ranging error and standard deviation curves of both are shown in Figure 14. The binocular ranging error is significantly reduced after the focal length correction. The average error changes from 0.184 m to 0.017 m, a reduction of 90.57%, which demonstrates the effectiveness of the laser rangefinder in assisting binocular camera calibration.

**Table 6.** Ranging results of binocular camera before and after focal length correction.

Distance	1.2 m	1.4 m	1.6 m	1.8 m	2.0 m	2.2 m
$Error_b$	−0.1239 m	−0.1531 m	−0.1729 m	−0.2130 m	−0.2379 m	−0.2894 m
$Error_a$	0.0164 m	0.0119 m	0.0055 m	−0.0121 m	−0.0292 m	−0.0489 m



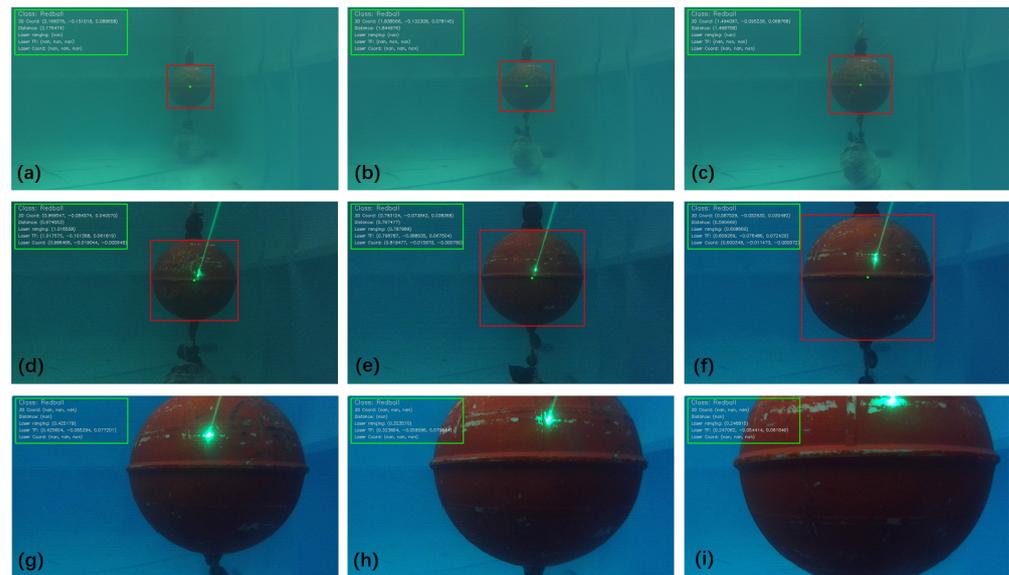
**Figure 14.** Error of binocular ranging before and after correction.

4.2. Experiment on Underwater Close-Range Large-Target Localization

In this study, the field of view of the binocular camera is filled only when the target with a 1 m diameter is 0.5 m away from the camera. To facilitate underwater experiments, a 0.34 m diameter floating ball is used as a substitute for the large target. We acquire floating ball images under different water quality and light conditions and train the network to obtain the weight model file after training with YOLOv8. Then, we deploy YOLOv8 on the NVIDIA Jetson NX of the FAUVMS platform to complete the preparation before target localization.

Experiments are conducted in the range of 0.2 m–2.2 m. The distance from the floating ball from 1.3 m to 2.2 m is set as the first stage, and the distance from the floating ball from 0.6 m to 1.2 m is set as the second stage. To prevent FAUVMS from colliding with the target while disposing of it, a safety distance of at least 0.2 m is maintained between the two, so the minimum distance of the method is set to 0.2 m, so the distance from the floating ball from 0.2 m to 0.5 m is set as the third stage. In the first stage, the binocular camera is used for distance estimation experiments at distances of 1.3 m, 1.4 m, 1.5 m, 1.6 m, 1.7 m, 1.8 m, 1.9 m, 2.0 m, and 2.1 m. In the second stage, the binocular camera and laser rangefinder are used for distance estimation experiments at 0.6 m, 0.7 m, 0.8 m, 0.9 m, 1.0 m, 1.1 m, and 1.2 m, and the fused result of the laser rangefinder and the binocular camera is considered the distance. In the third stage, assuming that the floating ball fills the field of view of the binocular camera and the binocular ranging fails, the laser rangefinder alone is used for the distance estimation experiment at 0.2 m, 0.3 m, 0.4 m, and 0.5 m. The three stages of floating ball localization are shown in Figure 15.

Since the floating ball is spherical, precisely placing the floating ball at the specified distance when it moves is difficult. The checkerboard is placed in front of the floating ball perpendicular to the ground and tangent to the floating ball. The thickness of the checkerboard is only 2 mm, so the checkerboard can be ignored. The distance between the binocular camera and the checkerboard is considered the distance between the binocular camera and the floating ball. In this way, the floating ball can be placed precisely at the specified distance. Each distance is measured eight times and the standard deviation is calculated and found to be relatively minimal. The errors between all averages and the real distance are also calculated. The experimental results of the relative distance estimation in the three stages are shown in Tables 7–9. A more intuitive comparison is shown in Figure 16.



**Figure 15.** The three stages of target localization. (a–c) represent the first stage at 2.1 m, 1.8 m, 1.5 m. (d–f) represent the second stage at 1.0 m, 0.8 m, 0.6 m. (g–i) represent the third stage at 0.4 m, 0.3 m, 0.2 m.

**Table 7.** The distance estimation results of the first stage.

Distance	1.4 m	1.5 m	1.6 m	1.7 m	1.8 m	1.9 m	2.0 m
$D_1$	1.3902 m	1.4941 m	1.6151 m	1.7286 m	1.8386 m	1.9309 m	2.0435 m
$Error_1$	0.0098 m	0.0059 m	0.0151 m	0.0286 m	0.0386 m	0.0309 m	0.0435 m

**Table 8.** The distance estimation results of the second stage.

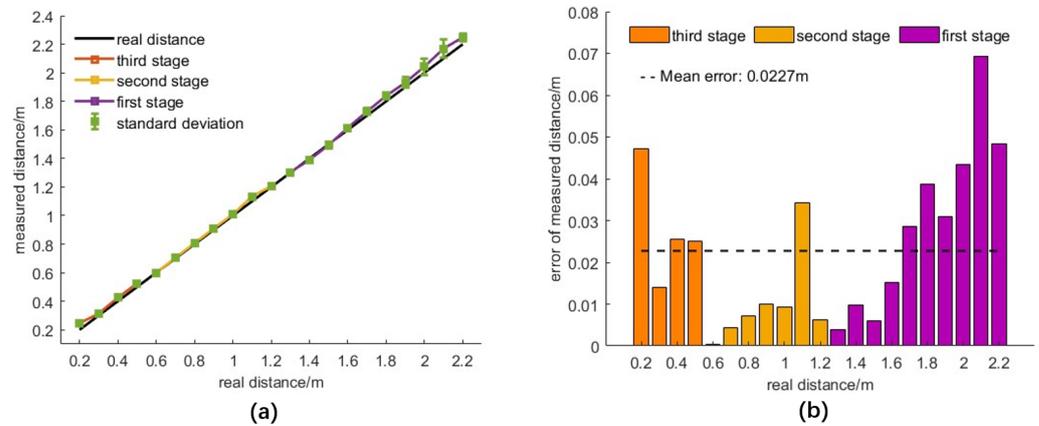
Distance	0.6 m	0.7 m	0.8 m	0.9 m	1.0 m	1.1 m	1.2 m
$D_2$	0.6003 m	0.7043 m	0.8071 m	0.9101 m	1.0092 m	1.1343 m	1.2063 m
$Error_2$	0.0003 m	0.0043 m	0.0071 m	0.0101 m	0.0092 m	0.0343 m	0.0063 m

**Table 9.** The distance estimation results of the third stage.

Distance	0.2 m	0.3 m	0.4 m	0.5 m
$D_3$	0.2471 m	0.3139 m	0.4256 m	0.5251 m
$Error_3$	0.0411 m	0.0139 m	0.0256 m	0.0251 m

To satisfy FAUVMS that it can intervene at the designated position of the target, the proposed method needs to achieve target localization accuracy within at least 10 cm. Figure 16 shows that the proposed method can meet the need and estimate the relative distance of the floating ball more accurately and stably in the three stages. As illustrated in Figure 16b, in the first stage, as the distance increases, the distance error of the binocular camera increases, but the maximum error is only approximately 7 cm. The error jump occurs because the distance measurement method of the binocular camera, in which disparity is used to calculate distance, is not very stable. In the second stage, errors are within 2 cm, excluding the error at 1.1 m, which reaches 3.4 cm. The outlier is caused by the instability of the binocular camera in recognizing the laser spot, but the error in the distance returned by the laser rangefinder is small. Compared with that in the second stage, the error in the third stage is higher, and an error jump occurs, but all the errors remain within 5 cm. The errors increase because the floating ball is round. As the distance decreases, the position of

the laser spot in the floating ball deviates, so the error becomes larger. In addition, the error jump is caused by protrusions on the floating ball surface. Overall, the average error of the three stages is only 2.27 cm. The proposed method can effectively achieve the localization of large underwater targets in close range.



**Figure 16.** (a) The distance estimation results of three stages. (b) The distance estimation errors of three stages.

## 5. Conclusions

In this paper, for underwater close-range large-target localization, we propose a multi-stage optical localization method that combines a binocular camera and a laser rangefinder. To improve the underwater ranging accuracy of the binocular camera, we modify the imaging model of the underwater camera to ensure that it continually fits the pinhole camera model, and we further improve the underwater calibration results of the binocular camera by using a laser rangefinder. To achieve accurate underwater target recognition, we adopt YOLOv8 to train the target image dataset from different pools and under different light conditions. The test results show that the AP (Average Precision) of the target reaches 0.982 when IoU (Intersection over Union) = 0.5. To align the coordinate systems of the binocular camera and laser rangefinder, we extrinsically calibrate the binocular camera and laser rangefinder. For target localization, we divide the localization process into three stages and the distances obtained by the binocular camera and laser rangefinder are fused via the KF. The pool experiment is conducted on the FAUVMS platform.

The experimental results show that the proposed method can work stably and effectively in the localization of large underwater close-range targets. The average distance error is only 2.27 cm, and the maximum error approaches 7 cm in the three stages. Compared with the method using a binocular camera alone, the proposed method does not exhibit localization failure when facing large targets at close ranges, and the error in underwater binocular localization is reduced by 90.57%.

In this study, the experiments are conducted in a static clear water pool and are not attempted in a turbulent underwater environment. In this environment, in addition to the effects of water and lighting conditions, the target image may be blurred and the laser rangefinder may work unstably, which impacts target localization. Therefore, in future work, we will further investigate the proposed method's capability to adapt to challenging water and light conditions as well as turbulent environments.

**Author Contributions:** Conceptualization, Q.Z. and Q.T.; methodology, W.X. and Q.T.; writing—original draft preparation, W.X.; writing—review and editing, W.X., X.Z. and Q.T.; funding acquisition, Q.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Youth Innovation Promotion Association, Chinese Academy of Sciences (2023208), and the fundamental research project of SIA.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data can be obtained from the corresponding author.

**Acknowledgments:** Thanks to all the reviewers for their contributions to improving the quality of this paper.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Zhou, Z.; Jiang, Y.; Li, Y.; Jian, C.; Sun, Y. A single acoustic beacon-based positioning method for underwater mobile recovery of an AUV. *Int. J. Adv. Robot. Syst.* **2018**, *15*, 1729881418801739. [CrossRef]
- Yan, Z.; Li, J.; Jiang, A.; Wang, L. An Obstacle Avoidance Algorithm for AUV Based on Obstacle's Detected Outline. In Proceedings of the 2018 37th Chinese Control Conference (CCC), Wuhan, China, 25–27 July 2018; pp. 5257–5262. [CrossRef]
- Ji-yong, L.; Hao, Z.; Hai, H.; Xu, Y.; Zhaoliang, W.; Lei, W. Design and Vision Based Autonomous Capture of Sea Organism With Absorptive Type Remotely Operated Vehicle. *IEEE Access* **2018**, *6*, 73871–73884. [CrossRef]
- Henson, B.T.; Zakharov, Y.V. Attitude-Trajectory Estimation for Forward-Looking Multibeam Sonar Based on Acoustic Image Registration. *IEEE J. Ocean. Eng.* **2019**, *44*, 753–766. [CrossRef]
- Lin, Y.; Hsiung, J.; Piersall, R.; White, C.; Lowe, C.G.; Clark, C.M. A Multi-Autonomous Underwater Vehicle System for Autonomous Tracking of Marine Life. *J. Field Robot.* **2017**, *34*, 757–774. [CrossRef]
- Cong, Y.; Gu, C.; Zhang, T.; Gao, Y. Underwater robot sensing technology: A survey. *Fundam. Res.* **2021**, *1*, 337–345. [CrossRef]
- Zhang, J.; Zhang, T.; Shin, H.S.; Wang, J.; Zhang, C. Geomagnetic Gradient-Assisted Evolutionary Algorithm for Long-Range Underwater Navigation. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–12. [CrossRef]
- Feezor, M.; Yates Sorrell, F.; Blankinship, P.; Bellingham, J. Autonomous underwater vehicle homing/docking via electromagnetic guidance. *IEEE J. Ocean. Eng.* **2001**, *26*, 515–521. [CrossRef]
- Wang, T.; Zhao, Q.; Yang, C. Visual navigation and docking for a planar type AUV docking and charging system. *Ocean. Eng.* **2021**, *224*, 108744. [CrossRef]
- He, Q.; Wang, Z.; Zeng, H.; Zeng, Y.; Liu, Y.; Liu, S.; Zeng, B. Stereo RGB and Deeper LIDAR-Based Network for 3D Object Detection in Autonomous Driving. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 152–162. [CrossRef]
- Zhang, Q.; Zhang, A.; Gong, P.; Quan, W. Research on Autonomous Grasping of an UVMS With Model-known Object Based On Monocular Visual System. In Proceedings of the ISOPE International Ocean and Polar Engineering Conference, Beijing, China, 20–25 June 2010; p. ISOPE-I-10–296. Available online: <https://onepetro.org/ISOPEIOPEC/proceedings-pdf/ISOPE10/All-ISOPE10/ISOPE-I-10-296/1714002/isope-i-10-296.pdf> (accessed on 20 June 2010).
- Xu, Z.; Haroutunian, M.; Murphy, A.J.; Neasham, J.; Norman, R. An Underwater Visual Navigation Method Based on Multiple ArUco Markers. *J. Mar. Sci. Eng.* **2021**, *9*, 1432. [CrossRef]
- Garrido-Jurado, S.; Muñoz-Salinas, R.; Madrid-Cuevas, F.; Marín-Jiménez, M. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognit.* **2014**, *47*, 2280–2292. [CrossRef]
- Chavez, A.G.; Mueller, C.A.; Doernbach, T.; Birk, A. Underwater navigation using visual markers in the context of intervention missions. *Int. J. Adv. Robot. Syst.* **2019**, *16*, 1729881419838967. [CrossRef]
- Li, Y.; Jiang, Y.; Cao, J.; Wang, B.; Li, Y. AUV docking experiments based on vision positioning using two cameras. *Ocean. Eng.* **2015**, *110*, 163–173. [CrossRef]
- Gao, X.S.; Hou, X.R.; Tang, J.; Cheng, H.F. Complete solution classification for the perspective-three-point problem. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 930–943. [CrossRef]
- Wang, Y.; Wang, S.; Wei, Q.; Tan, M.; Zhou, C.; Yu, J. Development of an Underwater Manipulator and Its Free-Floating Autonomous Operation. *IEEE/ASME Trans. Mechatronics* **2016**, *21*, 815–824. [CrossRef]
- Meng, Y.; Wu, Z.; Li, Y.; Chen, D.; Tan, M.; Yu, J. Vision-Based Underwater Target Following Control of an Agile Robotic Manta with Flexible Pectoral Fins. *IEEE Robot. Autom. Lett.* **2023**, *8*, 2277–2284. [CrossRef]
- Zhong, L.; Li, D.; Lin, M.; Lin, R.; Yang, C. A Fast Binocular Localisation Method for AUV Docking. *Sensors* **2019**, *19*, 1735. [CrossRef]
- Cain, C.; Leonessa, A. Laser based rangefinder for underwater applications. In Proceedings of the 2012 American Control Conference (ACC), Montreal, QC, Canada, 27–29 June 2012; pp. 6190–6195. [CrossRef]
- Utsumi, T.; Watanabe, K.; Nagai, I. A Range-finding System Using Multiple Lasers for an Underwater Robot with Pectoral-fin Propulsion Mechanisms and Improving Its Accuracy by a Gimbal Mechanism. In Proceedings of the 2021 IEEE International Conference on Mechatronics and Automation (ICMA), Takamatsu, Japan, 8–11 August 2021; pp. 681–686. [CrossRef]
- Bodenmann, A.; Thornton, B.; Ura, T. Generation of High-resolution Three-dimensional Reconstructions of the Seafloor in Color using a Single Camera and Structured Light. *J. Field Robot.* **2017**, *34*, 833–851. [CrossRef]
- Bleier, M.; van der Lucht, J.; Nüchter, A. SCOUT3D—An Underwater Laser Scanning System for Mobile Mapping. *ISPRS—Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *42*, 13–18. [CrossRef]
- Palomer, A.; Ridao, P.; Ribas, D. Inspection of an underwater structure using point-cloud SLAM with an AUV and a laser scanner. *J. Field Robot.* **2019**, *36*, 1333–1344. [CrossRef]

25. Castellón, M.; Palomer, A.; Forest, J.; Ridao, P. State of the Art of Underwater Active Optical 3D Scanners. *Sensors* **2019**, *19*, 5161. [[CrossRef](#)]
26. Hanai, A.; Choi, S.; Yuh, J. A new approach to a laser ranger for underwater robots. In Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No. 03CH37453), Las Vegas, NV, USA, 27–31 October 2003; Volume 1, pp. 824–829. [[CrossRef](#)]
27. Li, K.; Yang, S.; Liao, Y.Q.; Lin, X.T.; Wang, X.; Zhang, J.Y.; Li, Z. Underwater ranging with intensity modulated 532 nm laser source. *Acta Phys. Sin.* **2021**, *70*, 084203. [[CrossRef](#)]
28. Zheng, X.; Tian, Q.; Zhang, Q. Development and Control of an Innovative Underwater Vehicle Manipulator System. *J. Mar. Sci. Eng.* **2023**, *11*, 548. [[CrossRef](#)]
29. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [[CrossRef](#)]
30. Huang, H.; Zhou, H.; Qin, H.d.; Sheng, M.w. Underwater vehicle visual servo and target grasp control. In Proceedings of the 2016 IEEE International Conference on Robotics and Biomimetics (ROBIO), Qingdao, China, 3–7 December 2016; pp. 1619–1624. [[CrossRef](#)]
31. Zhou, Y.; Li, Q.; Ye, Q.; Yu, D.; Yu, Z.; Liu, Y. A binocular vision-based underwater object size measurement paradigm: Calibration-Detection-Measurement (C-D-M). *Measurement* **2023**, *216*, 112997. [[CrossRef](#)]
32. Treibitz, T.; Schechner, Y.Y.; Kunz, C.; Singh, H. Flat Refractive Geometry. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 51–65. [[CrossRef](#)]
33. Redmon, J.; Divvala, S.K.; Girshick, R.B.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 779–788.
34. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934v1.
35. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475. [[CrossRef](#)]
36. Reis, D.; Kupec, J.; Hong, J.; Daoudi, A. Real-Time Flying Object Detection with YOLOv8. *arXiv* **2023**, arXiv:2305.09972.
37. Liu, Z.; Lu, D.; Qian, W.; Gu, G.; Zhang, J.; Kong, X. Extrinsic calibration of a single-point laser rangefinder and single camera. *Opt. Quantum Electron.* **2019**, *51*, 1–13. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.