IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 29, 2021, accepted May 20, 2021, date of publication May 27, 2021, date of current version June 7, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3084153

An Object Localization System Using Monocular Camera and Two-Axis-Controlled Laser Ranging Sensor for Mobile Robot

HONGCHANG GE[1], TAO WANG[1,2,3,4], (Senior Member, IEEE), YUNCE ZHANG[1], AND SHIQIANG ZHU[1]

[1]Ocean College, Zhejiang University, Zhoushan 316000, China
[2]State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, Hangzhou 310007, China
[3]Engineering Research Center of Oceanic Sensing Technology and Equipment, Ministry of Education, Zhoushan 316000, China
[4]Key Laboratory of Ocean Observation-Imaging Testbed of Zhejiang Province, Zhoushan 316000, China

Corresponding author: Tao Wang (twang001@126.com)

This work was supported by the Research Project of Robotics Institute, Zhejiang University, under Grant K11804.

ABSTRACT Aiming at object grasping of indoor mobile robots, this paper develops an object localization system based on the fusion of a monocular camera and a two-axis-controlled laser ranging sensor. The system can locate the specified object in space without obtaining the background depth information of the object, which reduces the consumption of computational resources and has a low cost. According to the coordinate mapping of the object in the image, the measurement point of the laser ranging sensor is controlled to coincide with the target point by two servo motors in perpendicular axes. Thus, the spatial location of the target point can be determined from the rotation angles of the cradle head motors and the laser-measured distance together. To accurately control the rotation angles, a compound strategy combining feedback and feedforward is proposed. On the one hand, the feedback of the laser measurement point is introduced in the image coordinate to realize the close-loop control of the measurement point. On the other hand, a feedforward compensator based on disturbance observation is designed to reduce the influence of robot motion. An experimental setup is developed and substantial experiments are implemented to evaluate the performance of the proposed system. The results show that object localization can be effectively achieved with good accuracy and dynamic response.

INDEX TERMS Mobile robot, object localization, sensor fusion, laser ranging sensor, disturbance observation.

I. INTRODUCTION

Mobile robots can serve human beings in daily life and replace human beings to engage in dangerous and harsh work [1]–[3]. When mobile robots perform tasks such as detection and operation, they usually need to obtain the spatial location information of the object in the local coordinate of the robot. Therefore, real-time and accurate object localization technology is the premise to ensure the task completion of the robots.

At present, there have been a number of researches on object localization methods of mobile robots, including ultrasonic [4], Lidar [5], millimeter-wave radar [6], vision sensor [7] and other methods. Ultrasonic object localization is characterized by low cost and simple structure, but it is easily affected by ambient temperature [8]. It is often employed in the obstacle detection of the mobile robots [4]. Lidar object localization uses three-dimensional object segmentation method and directional enclosing box technol-ogy to separate point cloud objects, in which point cloud data directly contains the three-dimensional location information of the object [5]. Millimeter-wave radar is less affected by weather conditions and night, has a long detection range and is widely used outdoors. In addition to single sensor, object localization methods based on the fusion of radar and vision sensor have also been proposed in [6], [9], [10]. However, the point cloud data obtained by multi-line Lidar and millimeter-wave radar is relatively large and includes redundant background information, which requires more computational resources and hardware costs.

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Tan.

Computer vision includes monocular vision and depth vision. Monocular vision cannot directly obtain the three-dimensional space location of the object, and constraints such as known relative location of the camera [11], known shape and size of the object [12], and known ratio of distance to pixel area [7] must be added, leading to poor universality of this object localization method. Depth vision can be divided into binocular vision, multi-camera vision, structured light vision and time-of-flight vision according to different principles of depth information acquisition. Binocular vision imitates the perception principle of human eyes and uses two cameras with left and right distribution to simultaneously collect images and analyze the three-dimensional space location of the object. This method involves complex algorithms and requires good environment. At present, relevant studies mainly focus on feature matching and algorithm complexity reduction [13]–[15]. Multi-camera vision employs three or more cameras for object localization, to solve the problem of occlusion of the localization areas, but it requires complex computation. Structured light RGB-D cameras, such as Kinect v1 and Iphone X, adopt laser speckle coding principle to obtain depth information, which solves the problem of binocular feature matching. Time-of-flight RGB-D cameras, such as Kinect v2, utilize the time difference of optical signals to obtain depth data and have strong anti-interference ability. Since the depth image contains a lot of redundant background depth information, the measurement distance per unit power is relatively limited, and the data processing consumes a lot of computational resources.

Aiming at object grasping of indoor mobile robots, such as bottles, cups, apples, etc., it is only required to obtain the spatial location information of the object, and the background depth information is redundant. In order to meet the practical application requirements, the object localization method should have good dynamic performance and low cost. The object localization method based on single sensor is difficult to satisfy the requirement. Therefore, this paper proposes an object localization method based on the fusion of monocular camera and two-axis-controlled laser ranging sensor. Firstly, the image of the object is acquired by a monocular camera. Secondly, the coordinate of the target point in the image is determined by the existing object detection algorithm. After that, the measurement point of the laser ranging sensor is coincident with the target point by controlling the two-axis cradle head motors. Finally, the spatial location of the target point is solved jointly by integrating the rotation angles of the cradle head motors and the laser-measured distance. This method has the advantages of low cost, wide distance measuring range and good dynamic performance.

The contribution of this work is summarized as follows: 1) An object localization method based on fusion of monocular camera and two-axis-controlled laser ranging sensor is proposed, and the working principle of the method is described in detail; 2) A prototype is designed and assembled according to the object localization method; and 3) To realize the effective object localization of the mobile robot, a compound control strategy combining feedback and feedforward is developed. The effectiveness of the object localization method and the control strategy is verified by substantial experiments.

The rest of this paper is organized as follows. Relevant work is briefly reviewed in section II, the working principle and system design are discussed thoroughly in section III, the control strategy of the object localization system is expounded in detail in section IV, the experiments are demonstrated and analyzed in section V, and conclusions are drawn in section VI.

II. RELATED WORKS

Researchers have done a lot of research on different object localization methods based on single sensor. According to the known ratio of distance to pixel area, Bent *et al*. [7] used a monocular camera to locate the underwater object and completed the capture of the object by the under-water robot. Aim at solving the problem of picking cherry tomato, Kondo *et al*. [16] used binocular stereo vision technology to determine the three-dimensional position of fruit clusters and completed the automatic picking of fruits with a success rate of 70%. Bouazzaoui *et al*. [17] used an RGB-D camera to perceive and locate the surrounding environment of mobile robots, and the impact of the sensor acquisition modes on the localization accuracy was highlighted through a depth experimental research. Halterman *et al*. [18] used the Velodyne HDL-64E lidar to detect and locate obstacle on the water surface, and completed the obstacle avoidance of unmanned surface vehicle.

Although the object localization technology based on single sensor has realized its application value in different robot operating environments, each kind of sensor must have its own advantages and disadvantages, and different kinds of sensors have different advantages and disadvantages. It is difficult to achieve complementary advantages by using single sensor for object localization. Therefore, the object localization method based on fusion of multiple sensors is also proposed by researchers.

Broggi *et al*. [19] developed a pedestrian detection system aiming at localizing potentially dangerous situations under specific urban scenarios. This system uses the method based on fusion of laser scanner and vision system to search for pedestrians in key areas, which improves timewise performance and reduces false alarms. Nie *et al.* [20] proposed a multimodality fusion framework called Integrated Multimo-dality Fusion Deep Neural Network (IMF-DNN) for the multi-sensor information fusion of intelligent vehicle, which can flexibly accomplish the task of object detection and localization, and has superior robustness and generalization ability. In order to improve the navigation performance of mobile robots, Zhu *et al*. [21] proposed a rule-based reinforce-ment learning algorithm with space reduction. The obstacle environment perception of mobile robot is carried out by using lidar and ultrasonic sensors, and it is verified that the algorithm had good scalability by experiments.

Muresan *et al*. [22] used a loosely coupled sensor fusion approach to fuse the information obtained from trifocal camera, fisheye camera, long-range radar, 4-layer and 16-layer lidars. They realized the perception and localization of the objects on the road for autonomous vehicles.

At present, there have been a large number of researches on object localization methods based on fusion of monocular vision and laser, but most of them use monocular camera and multi-line lidar. The fused information contains a large amount of redundant background information which increa-ses the algorithm complexity and the consumption of computational resources, and leads to high cost. In this paper, the object localization method based on fusion of a monocular camera and two-axis-controlled laser ranging sensor solves this problem effectively. It is mainly suitable for mobile robots to grasp specific objects and other applications.

III. PRINCIPLE AND DESIGN

A. WORKING PRINCIPLE

Fig. 1 shows the structural principle of the object localization system. It mainly includes monocular camera, laser ranging sensor, cradle head pitching motor, cradle head horizontal motor and related structural parts. The object localization system obtains the image of the object by the monocular camera, selects the location of the object in the image through the frame of the existing object detection method, and selects the center point of the object as the target point which needs to be located. By controlling the rotation of the horizontal motor and the pitching motor of the two-axis cradle head, the yaw angle and pitching angle of the laser ray of the laser ranging sensor are adjusted to make the laser measurement point coincide with the desired target point, and the spatial location of the target point is solved jointly by integrating the rotation angles of the cradle head motors and the laser-measured distance. As shown in Fig. 1, O_C, O_J and O_Y represent camera optical center, laser optical center and laser light source respectively. Point O_J is located at the intersection of the central axis of the cradle head pitching motor, the central axis of the horizontal motor and the extension line of the laser light. When the cradle head motors rotate, the position of the laser optical center is fixed. In order to facilitate coordinate conversion and control of cradle head motors, the laser optical center is placed in the vertical downward direction of the camera optical center. Because of the offset distance between the laser optical center and the laser light source, the actual laser-measured distance is the sum of the O_JO_Y distance and the measured distance of the laser ranging sensor.

According to the image-forming principle of monocular camera [23], the target point in the world coordinate has only one coordinate point corresponding to it in the camera image-forming plane. Camera optical center, the target point in the image and the target point in the world coordinate are on the same straight line. Combined with the structural principle of the object localization system, the working principle as shown in Fig. 2 can be obtained. In Fig. 2, point A_1 is the

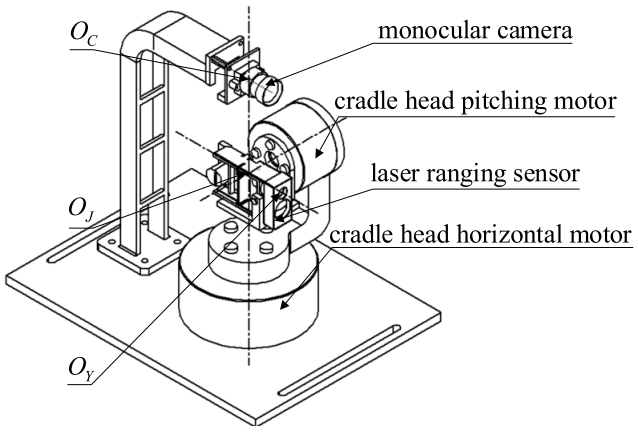


FIGURE 1. 3D structure of object localization system.

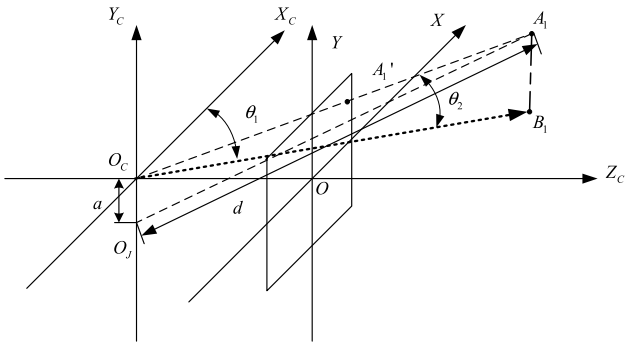


FIGURE 2. Working principle of object localization system.

target point which needs to be located, and point A_1' is the pixel coordinate point of the target point in the image. The image coordinate OXY and camera coordinate $O_CX_CY_CZ_C$ are respectively established with the image center point O and camera optical center O_C as the origin. The straight line O_JA_1 is the laser ray, and that point B_1 is the projection of the point A_1 in the plane $O_CX_CZ_C$. The angle θ_1 is the angle between the straight line O_CB_1 and X_C-axis. The angle θ_2 is the angle between the straight line O_CA_1 and O_CB_1. The d is the distance of O_JA_1, and the coordinate of laser optical center O_J in camera coordinate is $(0, -a, 0)$.

When the laser ray is parallel with the straight line O_CA_1, namely rotation angles of cradle head horizontal motor and pitching motor are θ_1 and θ_2 respectively, laser measurement point and the actual target point are in the same plane $O_CY_CB_1$, then the laser measurement point can be coincident with the actual target point only by adjusting the rotation angle of the cradle head pitching motor. It is assumed that the coordinate of the point A_1 in the camera coordinate is (x_c, y_c, z_c), and the coordinate of the point A_1' in the image coordinate is (x, y). When the laser measurement point coincides with point A_1, the rotation angle of cradle head pitching motor is θ_2'. Equations (1) and (2) can be described as

$$\begin{cases} \theta_1 = \dfrac{\pi}{2} - \arctan\left(\dfrac{xp}{f}\right) \\ \theta_2 = \arctan\left(\dfrac{yp}{f}\right) \end{cases} \tag{1}$$

where f is the focal distance of the camera, and p is the pixel size.

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} d\cos\theta_1\cos\theta_2' \\ d\sin\theta_2' - a \\ d\sin\theta_1\cos\theta_2' \end{bmatrix} \tag{2}$$

According to (1) and (2), the coordinate of the target point in the camera coordinate can be obtained.

B. SYSTEM DESIGN

The operating range of grasping daily objects of indoor mobile robots is required to be less than 10 meters, the accuracy of object localization is centimeter level, and the response frequency is 8 Hz. The selected parameters related to the laser ranging sensor and the monocular camera are shown in Table 1 below, where L is the measured distance of the laser ranging sensor.

TABLE 1. Parameters of laser ranging sensor and monocular camera.

Names of products	Parameters	
Laser ranging sensor	Laser wavelength	635 nm
	Overall dimensions	45 mm×25 mm×12 mm
	Measurement accuracy	±(1mm+L×0.004%)
	Measurement frequency	8 Hz
	Measurement range	0.03 m-40 m
Monocular camera	Pixel size	3 μm ×3 μm
	Overall dimensions	22.6 mm×22.6 mm×20 mm
	Maximum resolution	1920(H)×1080(V)
	Focal distance	3.6 mm
	Angle of view	90°

The main parameters of the cradle head motors is determined according to the dynamic requirement of the object localization system. According to the measurement frequency of the laser ranging sensor, the required adjustment time of position control of cradle head motor is chosen as 0.125 s, and the sampling frequency of the encoder is 200 Hz. In order to facilitate the calculation of the rated speed and rated torque of the cradle head motors, the following assumptions are made: 1) the mass distribution of the laser ranging sensor, cradle head motors and related structural parts is uniform and continuous; 2) ignore the influence of the cable of the laser ranging sensor and the cradle head motors.

For the cradle head pitching motor, the angular acceleration α is described as

$$\alpha = \frac{4\delta}{t^2} \tag{3}$$

where t is the adjustment time for the position control of the cradle head motor, and δ is the vertical angle of view of the monocular camera.

The rated speed n of the cradle head pitching motor is described as

$$n = \frac{\alpha t}{12} \tag{4}$$

According to (3) and (4), the required rated speed of the cradle head pitching motor is 129 r/min. Similarly, it can be obtained that the required rated speed of the cradle head horizontal motor is 206 r/min.

For the cradle head pitching motor, the moment of inertia J is described as

$$J = \iiint_{V_j} l_j^2 \rho_j dV_j + \iiint_{V_{jy}} l_{jy}^2 \rho_{jy} dV_{jy} \tag{5}$$

where ρ_j is the density of the laser ranging sensor, l_j is the distance between the mass element of the laser ranging sensor and the rotating shaft of the pitching motor, and V_j is the volume of the laser ranging sensor. The ρ_{jy} is the density of the laser ranging sensor connector, l_{jy} is the distance between the mass element of the laser ranging sensor connector and the rotating shaft of the pitching motor, and V_{jy} is the volume of the laser ranging sensor connector.

The rated torque M of the cradle head pitching motor is described as

$$M = J\alpha \tag{6}$$

According to (3), (5) and (6), the required rated torque of the cradle head pitching motor is 0.028 N.m. Similarly, it can be obtained that the required rated torque of the cradle head horizontal motor is 0.48 N.m. According to the required rated speed, rated torque and overall dimensions of the cradle head motors, relevant parameters of the pitching motor and the horizontal motor are selected as shown in Table 2 below.

TABLE 2. Parameters of pitching motor and horizontal motor.

Names of products	Parameters	
Pitching motor	Rated speed	520 r/min
	Rated torque	0.22 N.m
	Accuracy of encoder	0.02°
	Overall dimensions	39.6 mm×39.6 mm×33 mm
	Weight	120 g
	Frequency of communication	500 Hz
Horizontal motor	Rated speed	940 r/min
	Rated torque	0.61 N.m
	Accuracy of encoder	0.02°
	Overall dimensions	69 mm×69 mm×28.9 mm
	Weight	255 g
	Frequency of communication	500 Hz

IV. CONTROL STRATEGY

To make the laser measurement point and the target point overlap accurately in real time during the movement of

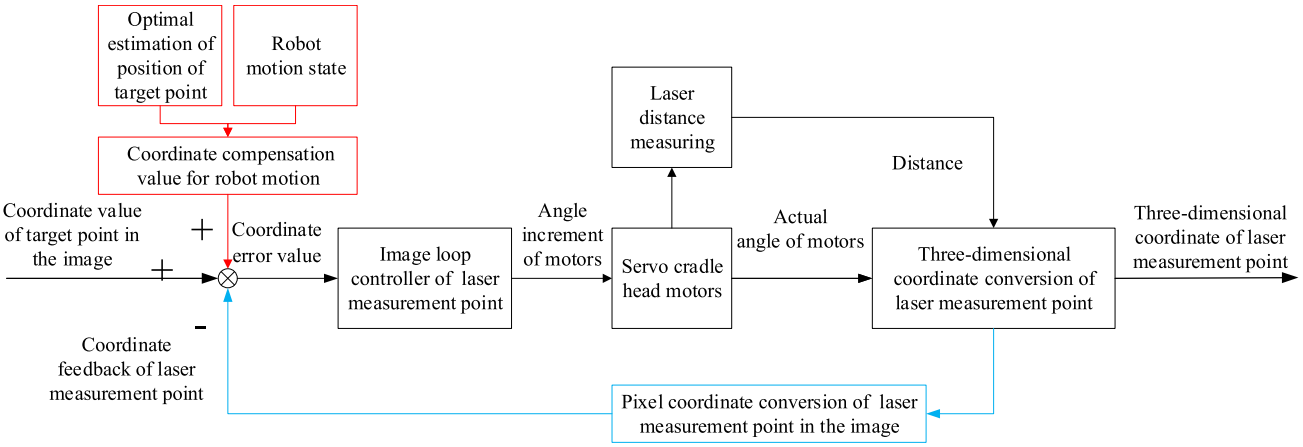


FIGURE 3. Block diagram of control strategy of object localization system using monocular camera and two-axis-controlled laser ranging sensor.

the mobile robot, the pixel coordinate feedback of the laser measurement point in the image is introduced to realize the close-loop control of the laser measurement point in the image coordinate. In addition, the influence of robot motion on object localization is analyzed, and the coordinate position error of laser measurement point is compensated through feedforward. The control block diagram of the object localization system is shown in Fig. 3.

A. CLOSED-LOOP CONTROL OF IMAGE COORDINATE POSITION OF LASER MEASUREMENT POINT

Due to the deviation of the actual installation position, the laser optical center and the camera optical center cannot be at the same point. When the laser ray is parallel with the connecting line between the target point and the camera optical center, the laser measurement point and the target point cannot overlap, namely it is impossible to make the laser measurement point and the target point overlap directly by controlling the two-axis cradle head motors. Therefore, the coordinate position feedback of laser measurement point in the image coordinate is introduced to carry out closed-loop feedback control in order to improve the accuracy of object localization.

Feedback information is usually obtained in two ways: direct feedback and indirect feedback. Generally speaking, the feedback information of pixel coordinate of laser measurement point is mainly obtained by image processing, but when the laser measurement point is far away and the illumination is strong, the recognition effect of laser measurement point in the image is poor, and it is difficult to be universal. In this paper, a new feedback method of pixel coordinate of laser measurement point in the image is proposed. The coordinate of laser measurement point in the camera coordinate is solved according to the rotation angles of the cradle head motors and the laser-measured distance, and the coordinate of laser measurement point in the image coordinate is reversely solved by coordinate transformation.

It is assumed that the rotation angles of the horizontal motor and the pitching motor of the cradle head are θ and φ, the coordinates of the laser measurement point are (x_1, y_1) in the image coordinate, and (x_{c1}, y_{c1}, z_{c1}) in the camera coordinate. Equations (7) and (8) can be obtained by combining (2) and the image-forming principle of camera.

$$\begin{bmatrix} x_{c1} \\ y_{c1} \\ z_{c1} \end{bmatrix} = \begin{bmatrix} d\cos\theta\cos\varphi \\ d\sin\varphi - a \\ d\sin\theta\cos\varphi \end{bmatrix} \tag{7}$$

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} \frac{x_{c1}f}{z_{c1}p} \\ \frac{y_{c1}f}{z_{c1}p} \end{bmatrix} \tag{8}$$

According to (7) and (8), the coordinate of the laser measurement point in the image coordinate can be obtained.

The closed-loop controller of image coordinate position of laser measurement point adopts proportional integral (PI) control, which has the characteristics of easy implementation, simple structure, convenient parameter adjustment, etc. It is widely used in the industrial control process [24], [25].The discretization difference equation of PI control of image coordinate position of laser measurement point is described as

$$u(k) = K_p(R(k) - H(k)) + K_i \sum_{i=0}^{k} (R(i) - H(i)) \tag{9}$$

where $u(k)$ is the angle increment value of the cradle head motors at k moment, K_p is the proportion coefficient, K_i is the integration coefficient, $R(k)$ is the image pixel coordinate value of the target point at k moment, and $H(k)$ is the image pixel coordinate value of the laser measurement point at k moment.

According to (7), (8) and (9), the closed-loop control of laser measurement point in image coordinate is illustrated in Algorithm. 1.

Algorithm 1 Closed-Loop Control of Image Coordinate Position of Laser Measurement Point

Input: angle of horizontal motor at k moment $\theta(k)$; angle of pitching motor at k moment $\varphi(k)$; actual laser-measured distance at k moment $d(k)$; image pixel coordinate of the target point at k moment $R(k)$

Output: coordinate of the laser measuring point in camera coordinate at k moment $J(k)$

```
Initialize k = 0
Initialize θ(k) = 0; φ(k) = 0
while pixel coordinate error between the laser measurement point and the target point is not 0 do
    Get J(k) using (7)
    Get H(k) using (8)
    Get u(k) using (9)
    Send position control instruction of servo cradle head motors
    k ← k + 1
    Update θ(k); φ(k); d(k)
End
```

B. COORDINATE POSITION ERROR COMPENSATION FOR ROBOT MOTION

If the laser measurement point is completely coincident with the target point at the current moment, the target point can be accurately located. However, the object localization system is mounted on the mobile robot. As the robot moves, the object localization system will move relative to the target point, and the coordinate position error between the laser measurement point and the target point in the image will increase. Since the movement state of the mobile robot can be directly measured by the sensors, the coordinate position error caused by the robot motion can be directly compensated by analyzing the influence of the robot motion on the object localization. Due to the fast speed of image processing to obtain the pixel coordinate position of the target point and short sampling time, the displacement movement of the robot within the sampling period can be approximated as uniform accelerated linear motion, and the attitude movement can be approximated as uniform rotation.

The world coordinate $O_W X_W Y_W Z_W$ is established with the target point as the origin O_W and the horizontal plane as the plane $X_W O_W Z_W$. The robot coordinate $O_R X_R Y_R Z_R$ is established with the geometric center of the mobile robot as the origin O_R and the transverse and longitudinal axes of the vehicle body as X_R-axis and Z_R-axis. The robot absolute coordinate $O_A X_A Y_A Z_A$ is established with the geometric center of the mobile robot as the origin O_A. Its X_A-axe, Y_A-axe and Z_A-axe are parallel to X_W-axe, Y_W-axe, Z_W-axe respectively. The working principle of coordinate position error compensation between the target point and the laser measurement point for robot motion is shown in Fig.4, and the angle θ_0 is the yaw angle of robot motion in the horizontal ground, namely that is the angle between the Z_R-axis and the Z_A-axis.

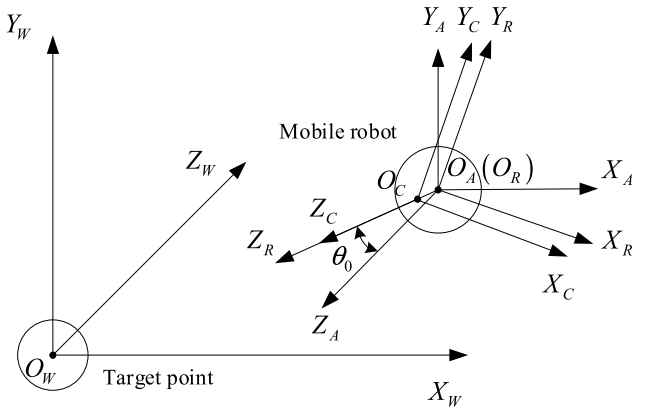


FIGURE 4. Working principle of coordinate position error compensation between the target point and the laser measurement point for robot motion.

It is assumed that at k moment, the coordinate position error between the target point and the laser measurement point is 0, the coordinates of the target point are (x_{ck}, y_{ck}, z_{ck}) in the camera coordinate, (x_{rk}, y_{rk}, z_{rk}) in the robot coordinate and (x_{ak}, y_{ak}, z_{ak}) in the robot absolute coordinate. The yaw angle is θ_{0k}. Equation (10) can be obtained from (7).

$$\begin{bmatrix} x_{ck} \\ y_{ck} \\ z_{ck} \end{bmatrix} = \begin{bmatrix} d\cos\theta\cos\varphi \\ d\sin\varphi - a \\ d\sin\theta\cos\varphi \end{bmatrix} \tag{10}$$

The monocular camera is fixed on the mobile robot, so the coordinate transformation relationship of the target point in the robot coordinate and the camera coordinate can be obtained through the rotation matrix R_{cr} and the translation vector T_{cr} as shown in (11).

$$\begin{bmatrix} x_{rk} \\ y_{rk} \\ z_{rk} \\ 1 \end{bmatrix} = \begin{bmatrix} R_{cr} & T_{cr} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{ck} \\ y_{ck} \\ z_{ck} \\ 1 \end{bmatrix} \tag{11}$$

In order to facilitate coordinate conversion, the monocular camera lens is mounted in the front of the mobile robot, then the camera coordinate only translates relative to the robot coordinate. Equation (12) is described as

$$R_{cr} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad T_{cr} = \begin{bmatrix} a_1 \\ b_1 \\ c_1 \end{bmatrix} \tag{12}$$

where (a_1, b_1, c_1) is the coordinate of the camera optical center in the robot coordinate.

The coordinate transformation relationship of the target point in the robot coordinate and the robot absolute coordinate can be described as

$$\begin{bmatrix} x_{ak} \\ y_{ak} \\ z_{ak} \\ 1 \end{bmatrix} = \begin{bmatrix} R_{ra} & T_{ra} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{rk} \\ y_{rk} \\ z_{rk} \\ 1 \end{bmatrix} \tag{13}$$

where

$$R_{ra} = \begin{bmatrix} \cos\theta_{0k} & 0 & \sin\theta_{0k} \\ 0 & 1 & 0 \\ -\sin\theta_{0k} & 0 & \cos\theta_{0k} \end{bmatrix}, \quad T_{ra} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

It is assumed that the coordinates of the center point of the robot in the world coordinate at k moment and $k+1$ moment are $(x_{w1k}, y_{w1k}, z_{w1k})$ and $(x_{w1k+1}, y_{w1k+1}, z_{w1k+1})$ respectively. According to the relative position of the object, Equation (14) can be obtained.

$$\begin{cases} (x_{w1k}, y_{w1k}, z_{w1k}) = \left(-x_{ak,} - y_{ak,} z_{ak}\right) \\ (x_{w1k+1}, y_{w1k+1}, z_{w1k+1}) = \left(-x_{ak+1,} - y_{ak+1,} z_{ak+1}\right) \end{cases} \tag{14}$$

The road surface on which indoor mobile robots drive is relatively flat, and the change of the Y_W-axis coordinate of the center point of the mobile robot in the world coordinate is small. It is assumed that the displacement of mobile robot changes only in the plane $X_W O_W Z_W$. IMU is fixed on the mobile robot, and the acceleration and angular velocity of the robot in its own coordinate are measured by the accelerometer and gyroscope in IMU. It is assumed that the velocity of the robot in its own coordinate at k moment is (v_{xk}, v_{zk}), the acceleration is (a_{xk}, a_{zk}), and the angular velocity is w_k.

Kalman filter method is an optimal estimation algorithm based on linear minimum variance estimation [26], which mainly consists of prediction and correction. The recursion formula of Kalman filter method is given as

$$\begin{cases} \hat{X}'_{k+1} = A\hat{X}_k + BU_k \\ {P_{k+1}}' = AP_kA^T + Q \\ K = \dfrac{{P_{k+1}}'H^T}{H{P_{k+1}}'H^T + R} \\ \hat{X}_{k+1} = \hat{X}'_{k+1} + K(z_{k+1} - H\hat{X}_k) \\ P_{k+1} = (I - KH){P_{k+1}}' \end{cases} \tag{15}$$

where A is the system state transition matrix, B is the input gain matrix, $\hat{X}_k$ is the system state matrix at k moment, $\hat{X}'_{k+1}$ is the system prediction state matrix at $k+1$ moment, U_k is the system input matrix, P_k is the system error covariance matrix at k moment, P'_{k+1} is the system prediction error covariance matrix at $k+1$ moment, Q is the process noise covariance matrix, K is the Kalman gain, H is the observation matrix, R is the measurement noise covariance matrix, z_{k+1} is the system observation value at $k+1$ moment, $\hat{X}_{k+1}$ is the optimal estimation matrix of the system state at $k+1$ moment, P_{k+1} is the optimal estimation matrix of the system error covariance at $k+1$ moment, and I is the identity matrix.

In the current system, $\hat{X}_k = [x_{w1k}, z_{w1k}, v_{xk}, v_{zk}, \theta_{0k}, \omega_k]^T$, $U_k = [a_{xk}, a_{zk}]^T$. According to the robot kinematics formula and Kalman filter recursive formula, the system state prediction equation can be described as follows (16), as shown at the bottom of the next page. where T_R is the sampling period.

Since the attitude movement of the robot within the sampling period can be approximately uniform rotation, Equation (17) can be obtained as follows

$$\begin{bmatrix} x_{rk+1} \\ y_{rk+1} \\ z_{rk+1} \\ 1 \end{bmatrix} = \begin{bmatrix} R'_{ar} & T'_{ar} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{ak+1} \\ y_{ak+1} \\ z_{ak+1} \\ 1 \end{bmatrix} \tag{17}$$

where

$$R'_{ar} = \begin{bmatrix} \cos(\theta_{0k} + \omega_k T_R) & 0 & -\sin(\theta_{0k} + \omega_k T_R) \\ 0 & 1 & 0 \\ \sin(\theta_{0k} + \omega_k T_R) & 0 & \cos(\theta_{0k} + \omega_k T_R) \end{bmatrix},$$

$$T'_{ar} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

According to (11), the coordinate of the target point in the camera coordinate is described as follows.

$$\begin{bmatrix} x_{ck+1} \\ y_{ck+1} \\ z_{ck+1} \\ 1 \end{bmatrix} = \begin{bmatrix} R_{cr} & T_{cr} \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} x_{rk+1} \\ y_{rk+1} \\ z_{rk+1} \\ 1 \end{bmatrix} \tag{18}$$

According to the modeling principle of the camera, Equation (19) can be obtained as follows

$$\begin{cases} \Delta e_x = \dfrac{x_{k+1} - x_k}{p} = \left(\dfrac{x_{ck+1}}{z_{ck+1}} - \dfrac{x_{ck}}{z_{ck}}\right)\dfrac{f}{p} \\ \Delta e_y = \dfrac{y_{k+1} - y_k}{p} = \left(\dfrac{y_{ck+1}}{z_{ck+1}} - \dfrac{y_{ck}}{z_{ck}}\right)\dfrac{f}{p} \end{cases} \tag{19}$$

where (x_k, y_k) is the coordinate of the target point in the image coordinate at k moment, (x_{k+1}, y_{k+1}) is the coordinate of the target point in the image coordinate at $k+1$ moment, Δe_x is the pixel error value of the target point in the X-axis of the image coordinate, and Δe_y is the pixel error value of the target point in the Y-axis of the image coordinate.

According to (10)-(19), the implementation details of the coordinate position error compensation for robot motion are illustrated in Alg. 2. $C_T(k)$ is the coordinate of the target point in the camera coordinate at k moment. ${C_T}'(k+1)$ is the prediction coordinate of the target point in the camera coordinate at $k+1$ moment. $R_T(k)$ is the coordinate of the target point in the robot absolute coordinate at k moment. ${R_T}'(k+1)$ is the prediction coordinate of the target point in the robot absolute coordinate at $k+1$ moment. $W_R(k)$ is the coordinate of the center point of the robot in the world coordinate at k moment. ${W_R}'(k+1)$ is the prediction coordinate of the center point of the robot in the world coordinate at $k+1$ moment.

V. EXPERIMENTS AND RESULTS

A. EXPERIMENTAL PLATFORM

The mobile robot used in the experiment is Scout2.0 version of AgileX Robotics, its maximum speed without load is 1.5 m/s and it can take a spin turn. The control unit of cradle head motors adopts STM32F4 development board. The STM32F4 development board is equipped with more than ten kinds of standard peripheral interfaces which can effectively meet the data transmission of sensors such as laser ranging

Algorithm 2 Coordinate Position Error Compensation for Robot Motion

Input: angle of horizontal motor at k moment $\theta(k)$; angle of pitching motor at k moment $\varphi(k)$; actual laser-measured distance at k moment $d(k)$; velocity of robot at k moment $v(k)$; angular velocity of robot at k moment $w(k)$; acceleration of robot at k moment $a(k)$; system error covariance matrix at k moment P_k

Output: pixel coordinate error compensation of the target point at k moment $\Delta e(k)$

Initialize $k = 0$
Initialize $C_T(k)$ using (10)
while the robot is moving **do**
  Get $R_T(k)$ using (11) and (13)
  Get $W_R(k)$ using (14)
  Get $W_R{}'(k+1)$ using (16)
  Get $R_T{}'(k+1)$ using (14)
  Get $C_T{}'(k+1)$ using (17) and (18)
  Get $\Delta e(k)$ using (19)
  $k \leftarrow k+1$
  Update $\theta(k)$, $\varphi(k)$, $d(k)$, $a(k)$
  Update v(k), $w(k)$, P_k, $C_T(k)$ using (10) and (15)
End

sensor, cradle head motors, inertial measurement unit, etc. It is also equipped with multiple 5V and 3.3V power supply interfaces to directly supply power to the sensors which reduces the number of the batteries. The image recognition and processing unit adopts Jetson Nano board. The Jetson Nano board has strong ability of image processing, and we used its GPU to realize the existing object detection algorithm. The STM32F4 development board communicates with the Jetson Nano board by a USB to serial-port line. Programming languages used on STM32F4 development board and Jetson Nano board are C language and C++ language respectively, and their integrated development environments are Keil μ Vision5 software and robot operating system respectively. Fig.5 shows the assembly picture of experimental platform for the object localization system.

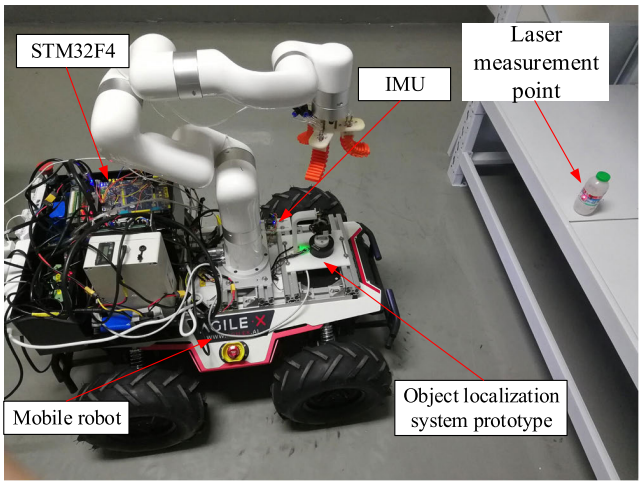


FIGURE 5. Assembly picture of experimental platform for the object localization system.

B. CLOSED-LOOP CONTROL PERFORMANCE

Single shot multibox detector (SSD) is an object detection algorithm which has high accuracy and fast processing speed [27]. In this paper, we used SSD algorithm to detect specific object in the image, and the target point needed to be located in the object is selected and marked with a blue circle in the image. The experimental results are shown in Fig. 6 by comparing the target point and the actual laser measurement point, in which the red and white spot is the actual laser measurement point.

The closed-loop control of image coordinate position of laser measurement point adopts PI controller. In order to evaluate its control performance, the pixel coordinate error between the target point and the laser measurement point in the image is selected as the quantitative index. The pixel size of the image plane used is 640 $\times$ 360, and the center of image is determined as the initial target point, namely the coordinate of initial target point is (320, 180). The coordinate of terminate target point is (220, 280), and the graph of pixel coordinate error between the target point and the laser measurement point in the image is shown in Fig. 7.

$$\begin{bmatrix} \hat{x}'_{w1k+1} \\ \hat{z}'_{w1k+1} \\ \hat{v}'_{xk+1} \\ \hat{v}'_{zk+1} \\ \hat{\theta}'_{0k+1} \\ \hat{w}'_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cos\theta_{0k}T_R & -\sin\theta_{0k}T_R & 0 & 0 \\ 0 & 1 & -\sin\theta_{0k}T_R & -\cos\theta_{0k}T_R & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & T_R \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{w1k} \\ z_{w1k} \\ v_{xk} \\ v_{zk} \\ \theta_{0k} \\ w_k \end{bmatrix} + \begin{bmatrix} 0.5\cos\theta_{0k}T_R^2 & -0.5\sin\theta_{0k}T_R^2 \\ -0.5\sin\theta_{0k}T_R^2 & -0.5\cos\theta_{0k}T_R^2 \\ T & 0 \\ 0 & T \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a_{xk} \\ a_{zk} \end{bmatrix} \tag{16}$$

FIGURE 6. Comparison of the selected target point and the laser measurement point in the image.

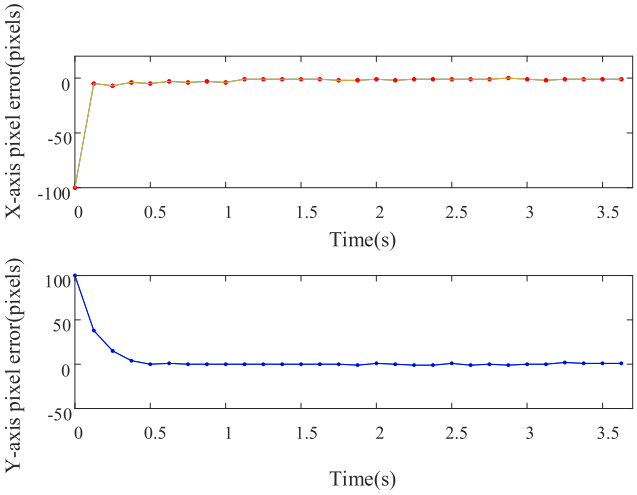


FIGURE 7. Pixel coordinate error between the target point and the laser measurement point in the X-axis and Y-axis directions of the image coordinate.

The pixel coordinate steady-state error between the target point and the laser measurement point in the X-axis and Y-axis directions of the image coordinate are both within the interval of [−2,2]. This method can well match the target point and the laser measurement point, and it has a high control performance.

The running time of the whole process from obtaining the image of the object to completing the accurate object localization is 0.6 s through actual measurement.

C. ACCURACY EVALUATION

In order to evaluate the location accuracy of this object localization system, the target points at different positions in the camera coordinate are located. By comparing the given standard value of the three-dimensional position of the target point with the actual measured value, the location error of the object localization system prototype is obtained as shown in Table 3 below. It can be seen from Table 3 that the location accuracy of object localization system prototype designed in this paper reaches the centimeter level, which can meet the requirements of indoor mobile robots grasping daily objects and other operations.

TABLE 3. Location error of object localization system prototype.

Number	Coordinate	x(mm)	y(mm)	z(mm)
1	Given standard value	0	0	1038
	Actual measured value	1	0	1046
	Location error	-1	0	-8
2	Given standard value	0	0	2061
	Actual measured value	1	3	2068
	Location error	-1	-3	-7
3	Given standard value	0	0	3046
	Actual measured value	3	-2	3051
	Location error	-3	2	-5
4	Given standard value	0	0	3997
	Actual measured value	3	2	4005
	Location error	-3	-2	-8
5	Given standard value	0	0	5025
	Actual measured value	3	-5	5031
	Location error	-3	5	-6
6	Given standard value	0	0	6014
	Actual measured value	4	4	6020
	Location error	-4	-4	-6
7	Given standard value	0	0	6988
	Actual measured value	8	-9	6997
	Location error	-8	9	-9

D. COORDINATE POSITION ERROR COMPENSATION

Fig. 8 shows the experimental principle of coordinate position error compensation for robot motion. A stationary target point is selected as the origin and the world coordinate is established. When the mobile robot is stationary, the laser measurement point of the object localization system mounted on the mobile robot can fall on the target point. To test the performance in dynamic situation, the mobile robot is controlled to rotate with a constant speed of 3°/s from the same position and posture under two conditions, with and without compensation. The pixel coordinate errors between the laser measurement point and the target point in the X-axis and Y-axis directions of the image coordinate are recorded when the system runs steadily, and the change curves of pixel coordinate error are shown in Fig. 9 and Fig. 10 respectively.

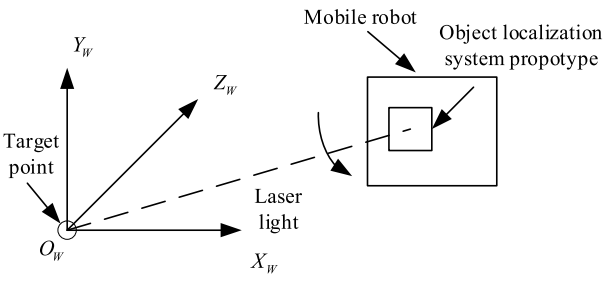


FIGURE 8. Experimental principle of coordinate position error compensation for robot motion.

The pixel coordinate errors between the laser measurement point and the target point in the X-axis and Y-axis directions of the image coordinate are affected by external interference, such as image pixel coordinate error of target point due to target detection, speed control precision of robot

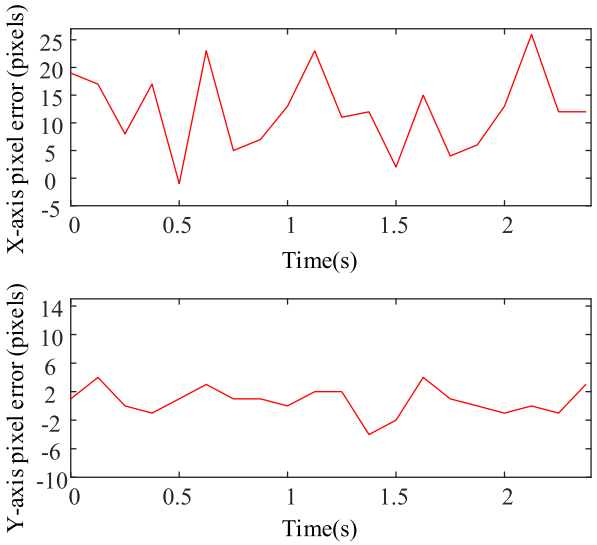


FIGURE 9. Pixel coordinate error between the laser measurement point and the target point in the X-axis and Y-axis directions of the image coordinate without compensation in dynamic situation.

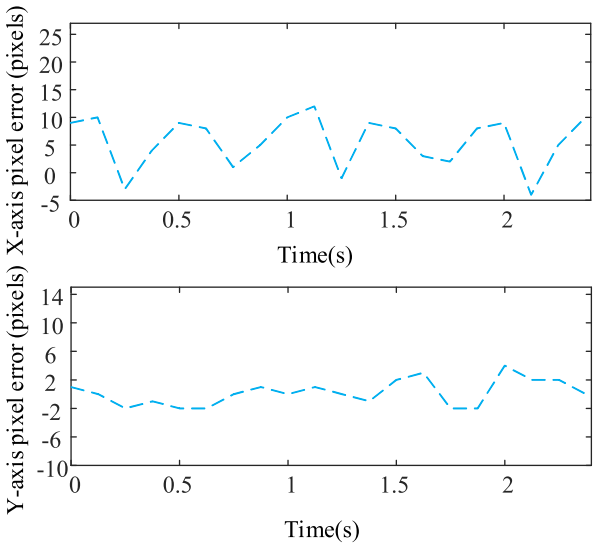
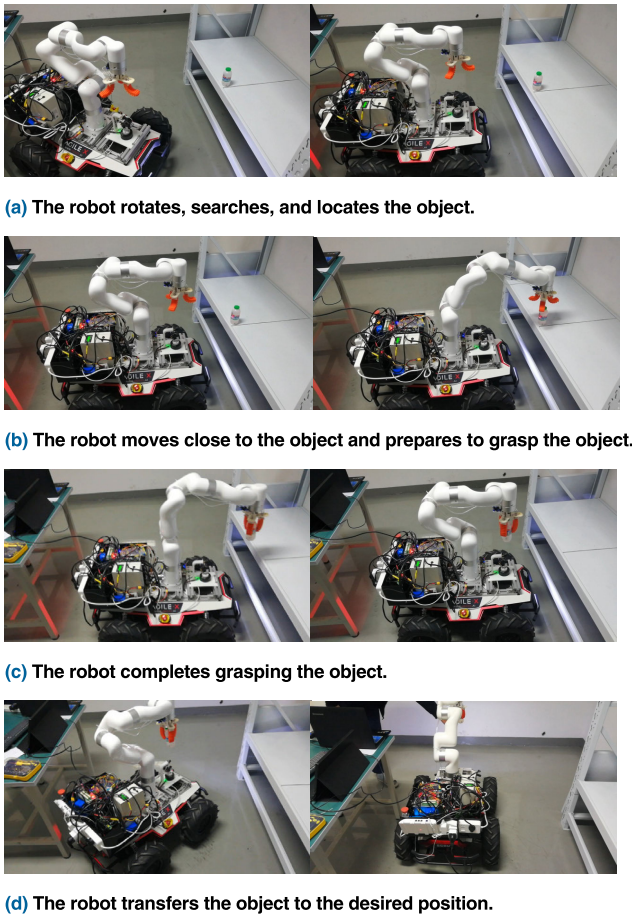


FIGURE 10. Pixel coordinate error between the laser measurement point and the target point in the X-axis and Y-axis directions of the image coordinate with compensation in dynamic situation.

motion, so there is fluctuation in the pixel coordinate error curve. However, for the trend of overall changes, the pixel coordinate error in the X-axis direction with compensation is closer to zero, and the pixel coordinate error in the Y-axis direction with compensation is basically within the range of the steady-state error $[-2,2]$. By comparing Fig. 9 and Fig. 10, it is proved that the effect of object localization with coordinate error compensation is better than that without coordinate error compensation, which verifies the validity of the coordinate error compensation theory when the robot is moving.

E. APPLICATION IN ROBOT GRASPING

Finally, the object localization system is applied in mobile robots to grasp daily objects. The object localization system prototype is installed on the mobile robot. The mobile robot searches the object within the operating range by its own motion. When the object is captured by the monocular camera, the three-dimensional location information of the object is provided to the mobile robot through the object localization system, which guides the robot to move close to the object. When the mobile robot reaches the specified position, the robot arm and soft claw are controlled to grasp the object, and the mobile robot transfers it to the desired position. Fig. 11 shows the experimental results of object grasping.



(a) The robot rotates, searches, and locates the object.

(b) The robot moves close to the object and prepares to grasp the object.

(c) The robot completes grasping the object.

(d) The robot transfers the object to the desired position.

FIGURE 11. Experimental results of object grasping by the mobile robot using the object localization system.

VI. CONCLUSION

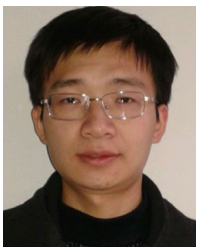
In this paper, an object localization method based on fusion of a monocular camera and a two-axis-controlled laser ranging sensor is developed. The working principle of the method is described in detail and a prototype of the object localization system is designed. The prototype can be applied to the actual needs of grasping daily objects for indoor mobile robot, with good dynamic localization performance and low cost. It can locate the selected object without obtaining the background depth information of the whole object, which reduces the consumption of computational resources. In order to improve

dynamic performance and accuracy, this work designs the closed-loop controller of laser measurement point in image coordinate and the disturbance compensator for robot motion. An experimental platform is built and a large number of experiments are carried out. The results show that the proposed system has good feasibility. In the future, the object localization system will be combined with object segmentation technology to achieve more interesting function.

REFERENCES

[1] D. Fischinger, P. Einramhof, K. Papoutsakis, W. Wohlkinger, P. Mayer, P. Panek, S. Hofmann, T. Koertner, A. Weiss, A. Argyros, and M. Vincze, "Hobbit, a care robot supporting independent living at home: First prototype and lessons learned," *Robot. Auto. Syst.*, vol. 75, pp. 60–78, Jan. 2016.
[2] D. Ball, B. Upcroft, G. Wyeth, P. Corke, A. English, P. Ross, T. Patten, R. Fitch, S. Sukkarieh, and A. Bate, "Vision-based obstacle detection and navigation for an agricultural robot," *J. Field Robot.*, vol. 33, no. 8, pp. 1107–1130, Dec. 2016.
[3] W. Wang, W. Dong, Y. Su, D. Wu, and Z. Du, "Development of search-and-rescue robots for underground coal mine applications," *J. Field Robot.*, vol. 31, no. 3, pp. 386–407, May 2014.
[4] Q. Huang, B. Hong, K. Javaid, Q. Gao, Y. Zhu, and Y. Ruan, "Ultrasonic location and obstacle avoidance device for autonomous soccer robot," *J. Harbin Inst. Technol.*, vol. 35, no. 9, pp. 1077–1079, Sep. 2003.
[5] D. C. Hoang, L. C. Chen, and T. H. Nguyen, "Sub-OBB based object recognition and localization algorithm using range images," *Meas. Sci. Technol.*, vol. 28, no. 2, pp. 25401–25411, Dec. 2016.
[6] W. Zhao, B. Li, Y. Shan, and H. Xu, "Vehicle detection and tracking based on fusion of millimeter wave radar and monocular vision," *Geomat. Inf. Sci. Wuhan Univ.*, vol. 44, no. 12, pp. 1832–1840, Dec. 2019.
[7] O. A. H. Bent, B. S. Martin, and S. Ingrid, "Monocular vision-based gripping of objects," *Robot. Auton. Syst.*, vol. 131, pp. 103589–103600, Sep. 2020.
[8] F. Tong, S. Xu, and T. Xu, "A high precision ultrasonic ranging processing method," *J. Xiamen Univ. (Natural Sci.)*, vol. 37, no. 4, pp. 507–512, Feb. 1998.
[9] Y. He, H. Jiang, H. Fang, Y. Wang, and Y. Liu, "Research progress of intelligent obstacle detection methods of vehicles and their application on agriculture," *Trans. Chin. Soc. Agric. Eng.*, vol. 34, no. 9, pp. 21–32, Jun. 2018.
[10] M. Perrollaz, R. Labayrade, C. Royere, N. Hautiere, and D. Aubert, "Long range obstacle detection using laser scanner and stereovision," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2006, pp. 182–187.
[11] H. Ji, M. Wei, H. Li, X. Li, and K. Zhang, "Method and virtual test of monocular distance measurement based on PreScan platform," *Meas. Control Technol.*, vol. 32, no. 12, pp. 49–56, Dec. 2013.
[12] D. Xiao and J. Zhai, "Target distance measurement method with monocular vision for wheeled mobile robot," *Comput. Eng.*, vol. 34, no. 4, pp. 287–291, Apr. 2017.
[13] D. Wang, W. Wang, and Q. Cheng, "Object recognition and localization based on binocular stereo vision," *J. Jilin Univ. (Inf. Sci. Ed.)*, vol. 34, no. 2, pp. 289–295, Mar. 2016.
[14] Y. Tuo, F. Shi, and T. Lin, "A method for obstacle detection and location based on mobile welding robot," *J. Shanghai Jiaotong Univ.*, vol. 44, pp. 126–129, Oct. 2010.
[15] X. J. Zou, H. X. Zou, and J. Lu, "Virtual manipulator-based binocular stereo vision positioning system and errors modelling," *Mach. Vis. Appl.*, vol. 23, no. 1, pp. 43–63, Jan. 2012.
[16] N. Kondo, Y. Nishitsuji, P. P. Ling, and K. C. Ting, "Visual feedback guided robotic cherry tomato harvesting," *Trans. ASAE*, vol. 39, no. 6, pp. 2331–2338, Feb. 1996.
[17] I. E. Bouazzaoui, S. R. Florez, and A. E. Ouardi, "Enhancing RGB-D SLAM performances considering sensor specifications for indoor localization," *IEEE Sensors J.*, early access, Apr. 16, 2021, doi: 10.1109/JSEN.2021.3073676.
[18] R. Halterman and M. Bruch, "Velodyne HDL-64E lidar for unmanned surface vehicle obstacle detection," *Proc. SPIE*, vol. 7692, Apr. 2010, Art. no. 76920D.
[19] A. Broggi, P. Cerri, S. Ghidoni, P. Grisleri, and H. Gi Jung, "A new approach to urban pedestrian detection for automatic braking," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 4, pp. 594–605, Dec. 2009.
[20] J. Nie, J. Yan, H. Yin, L. Ren, and Q. Meng, "A multimodality fusion deep neural network and safety test strategy for intelligent vehicles," *IEEE Trans. Intell. Vehicles*, vol. 6, no. 2, pp. 310–322, Jun. 2021, doi: 10.1109/TIV.2020.3027319.
[21] Y. Zhu, Z. Wang, C. Chen, and D. Dong, "Rule-based reinforcement learning for efficient robot navigation with space reduction," *IEEE/ASME Trans. Mechatronics*, early access, Apr. 13, 2021, doi: 10.1109/TMECH.2021.3072675.
[22] M. P. Muresan, I. Giosan, and S. Nedevschi, "Stabilization and validation of 3D object position using multimodal sensor fusion and semantic segmentation," *Sensors*, vol. 20, no. 4, pp. 1110–1142, Feb. 2020.
[23] X. Du, G. Xu, Y. Zhu, and Y. Wang, "A simple 3D reconstruction method based on single camera," *Chin. J. Sens. Actuators*, vol. 20, no. 8, pp. 1917–1920, Sep. 2007.
[24] L. Zhang and Y. Su, "Feedback control design of servo systems: A review," *Control Theor. Appl.*, vol. 31, no. 5, pp. 545–559, Jul. 2014.
[25] K. H. Ang, G. Chong, and Y. Li, "PID control system analysis, design, and technology," *IEEE Trans. Control Syst. Technol.*, vol. 13, no. 4, pp. 559–576, Jul. 2005.
[26] J. Pentzer, S. Brenman, and K. Reichard, "Model-based prediction of skid-steer robot kinematics using online estimation of track instantaneous centers of rotation," *J. Field Robot.*, vol. 31, no. 3, pp. 455–476, May 2014.
[27] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. ECCV*, Amsterdam, The Netherlands, Oct. 2016, pp. 21–37.

HONGCHANG GE received the B.S. degree in agricultural engineering from Nanjing Agricultural University, Nanjing, China, in 2018. He is currently pursuing the degree with the Ocean College, Zhejiang University, Zhoushan, China. His research interest includes mobile robots.



TAO WANG (Senior Member, IEEE) received the B.S. and Ph.D. degrees in mechanical engineering from Zhejiang University, Hangzhou, China, in 2008 and 2013, respectively. He is currently an Associate Professor with the Ocean College, Zhejiang University, Zhoushan, China. His research interests include mechatronic systems and robotics.

YUNCE ZHANG received the B.S. degree in mechanical engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2016. He is currently pursuing the Ph.D. degree with the Ocean College, Zhejiang University, Zhoushan, China. His research interest includes autonomous robots.

SHIQIANG ZHU received the B.S. and Ph.D. degrees in mechanical engineering from Zhejiang University, Hangzhou, China, in 1988 and 1995, respectively, and the M.S. degree in mechatronic engineering from the Beijing Institute of Technology, Beijing, China, in 1991. He has been a Faculty Member at Zhejiang University, since 1995, and was promoted to the rank of Professor, in 2001. He is the Director of the Zhejiang Laboratory. His research interests include robotics and mechatronics.