Research on Depth-Adaptive Dual-Arm Collaborative Grasping Method

Hao Zhang[1], Pengfei Yi[1], Rui Liu[1], Jing Dong[1], Qiang Zhang[1,2], and Dongsheng Zhou[1,2(✉)]

[1] The Key Laboratory of Advanced Design and Intelligent Computing, Ministry of Education, School of Software Engineering, Dalian University, Dalian, People's Republic of China
zhouds@dlu.edu.cn

[2] School of Computer Scicence and Technoloy, Dalian University of Technology, Dalian, People's Republic of China

Abstract. Among the existing dual-arm cooperative grasping methods, the dual-arm cooperative grasping method based on RGB camera is the mainstream intelligent method. However, these methods often require predefined depth, difficult to adapt to changes in depth without modification. To solve this problem, this paper proposes a dual-arm cooperative grasping method based on RGB camera, which is suitable for scenes with variable depth, to increase the adaptability of dual-arm cooperation. Firstly, we build a mathematical model based on RGB camera, and use the markers attached to the target to obtain the depth information of the target. Then the 3D pose of the target under the robot world coordinate system is obtained by combining the depth information and pixel information. Finally, the task is assigned to the left and right robotic arms, and the target grabbing task is realized based on the main-auxiliary control. The proposed approach is validated in multiple experiments on a Baxter robot under different conditions.

Keywords: Dual-arm collaboration · Target localization · Robotics

1 Introduction

The dual-arm collaboration [15,21] is an important part of the robotics field. Since some tasks will exceed the capabilities of a single robotic arm, dual-arm are required to cooperate and can be used in various tasks. For example, it is used to automated minimally invasive suturing [27], sorting of surgical instruments [23], harvesting of aubergine [19], as well as performing high-precision tasks [20]. It can also perform grasping and placing tasks that are laborious for human workers, non-fixed-point grasp to achieve autonomous assembly tasks and improve work efficiency by working with dual-arm.

Existing dual-arm collaboration methods still have some limitations. Using predefined paths and manual intervention methods to complete collaborative

H. Gao et al. (Eds.): CollaborateCom 2022, LNICST 461, pp. 268–285, 2022.
https://doi.org/10.1007/978-3-031-24386-8_15

tasks with a fixed target pose or by manipulating the arms in real time is not very intelligent. With the development of computer vision, a vision-guided dual-arm collaborative approach has emerged. Compared with the above methods, this method has higher intelligence and can estimate the pose information of the target through visual information when the depth information is known. Most of these methods use a fixed depth scene. However, when the depth is not fixed, these methods often resort to additional depth cameras to acquire the target's pose. This approach increases the cost of use while making deployment and integration more difficult.

Aiming at the limitation of the dual-arm cooperative method when the depth is variable, in order to improve the adaptability of the dual-arm cooperative grasping method based on the RGB camera, this paper proposes a dual-arm cooperative grasping method that only relies on the RGB camera. This method can be applied to scenes with varying depths. Depth information is lost when only RGB cameras are used, and depth information is difficult to obtain due to the lack of a known standard to measure depth changes. Inspired by the single-arm approach to grasping objects, we use mathematical modeling to obtain a deep informative model. Then the depth value is obtained from the depth information model and combined with the image information to estimate the position and pose information of the target. Finally, the target is grasped through the cooperation of the dual-arm. The main work of this paper is as follows:

- In this paper, we propose a dual-arm collaborative grasp method based on RGB cameras to obtain depth information through the established mathematical model, without relying on additional devices such as depth cameras.
- We propose a target position localization method based on a monocular RGB camera and a method to estimate the target pose from image information.
- We designed a dual-arm collaborative control strategy based on main-auxiliary control. It can steadily control the dual-arm to collaboratively grasp the target object.

2 Related Work

In industrial production, dual-arm achieve collaborative tasks by means of predefined paths, such as completing the board burr removal [16]. The advantage of this method is that it is more efficient, it can achieve interruption-free work, and can meet the production requirements, which plays an important role in industrial production. However, the robot performs repetitive operations, the arms cannot be flexibly adjusted. In addition, dual-arm collaboration methods are broadly classified into human intervention-based dual-arm collaboration and computer vision-based dual-arm collaboration. In the following, we will first introduce the human intervention-based and computer vision-based methods. Since there are many common problems in single-arm grasping and dual-arm grasping, the idea of single-arm grasping method to solve the problem is very worthy of reference. Then we will introduce the methods related to single-arm grasping. The introductions are respectively as follows.

2.1 Dual-Arm Collaboration Based on Human Intervention

Liang et al. [9] proposed a bilateral teleoperating system, where the operator can control the 6DOF pose and gripper width of each arm through an intuitive bilateral manipulator. Using the double-sided manipulator, the operation of the robotic arms on both sides is independent and does not affect each other, which is convenient for control. Tung et al. [22] proposed a method for remotely controlling robotic arm collaboration. Multiple users in different locations use mobile phones to control the robot arms. Since the delay of each user's network connection is different, it is adopted to wait for all mobile phone connections to receive new mobile phone information, and then start all robotic arms to ensure the coordination of multiple robotic arms. Laghi et al. [8] presented an alternative method for remotely manipulating robots. Not only independent control can be achieved, but also shared control can be achieved. Lipton et al. [10] proposed a method to control a dual-arm robot through a VR device, and the user could control the robot remotely. Bai et al. [1] presented a strategy of the dual-arm coordinated control for twisting manipulation. One of the robotic arms is controlled through predefined planning. The operator observes through the camera and controls the other arm in real time via teleoperation to complete the dual-arm collaboration.

Yu et al. [25] proposed a cooperative control strategy. The operator operates the master arm movement and returns the movement information of the master arm to the control terminal, and the control terminal controls the slave robot arm movement based on this information to achieve the effect that the slave arm follows the master arm movement. Ibarguren et al. [6] proposed a trajectory-driven cooperative task execution architecture. When the operator guides the robot, it follows a given trajectory to move smoothly. And impedance control is added to allow the operator to adjust the path.

2.2 Dual-Arm Collaboration Based on Computer Vision

Rastegarpanah et al. [17] described the realization of different grasping tasks based on computer vision. Firstly, by controlling the motion of the main arm, the double-arm operation is realized based on the method of motion tracking. The target is then grasped using the dual arms cooperatively based on a fixed depth and moved along a defined path. Medjram et al. [13] proposed a vision-based method for estimating the pose of a carton using edge features and perspective methods. Grasp point information is obtained by using a depth camera and modeling perspective changes. Zahavi et al. [26] presented use of dual-arm to perform pick and place tasks while using a single overhead camera system. It is used to detect and grasp the areas on both sides of the industrial robot, and classify it and place them in a specified location. The implemented machine visual algorithm can identify the color and shape of the target.

2.3 Single-Arm Based Grasping Methods

Yang et al. [24] presented a single-arm target grasping method, which can quickly recognize the attribute information of the object and complete the grasping. Griffin et al. [5] designed a framework to extract the pixel information of the target through video object segmentation technology, and calculate the depth value of the target through RGB camera motion, and then perform a single-arm grasp operation on the target. Lundell et al. [12] presented a top grasp planning method, in dense sampling of the scene, the grasp directions are selected from directly above the target and from five oblique angle positions, which utilize markers for position calibration. Du et al. [3] summarized the robot grasping method based on computer vision. It is divided into three parts, including the target positioning method, the target attitude estimation method, and the target grasping estimation method. Cai et al. [2] presented a grasp training system. The system can select the correct grab target according to the defined correction strategy.

In summary, the method based on predefined paths has stricter requirements for target position and lower ability to adapt to changes. The human intervention-based method has a high operator dependence, when operators become insufficient concentration during prolonged operations, this can lead to unnecessary hazards. The dual-arm collaborative method based on RGB camera has high intelligence, but there is difficulty in handling depth changes. To improve this situation, inspired by the single-arm grasping target methods. Firstly, we estimate the depth information based on the RGB camera using visual information combined with markers of known width. And the estimated depth information is calibrated by the established depth information model to obtain a more accurate depth value. Then the position and pose of the target are calculated based on the depth information. Finally, a dual-arm collaborative grasping of the target object is achieved based on the main-auxiliary control strategies.

3 Methods

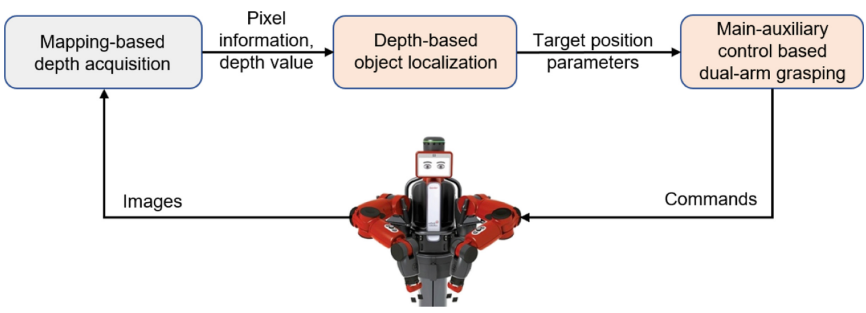

Fig. 1. Overview of the method.

This paper proposes a dual-arm cooperative grasping method suitable for depth-variable scenes. Based on the RGB camera, the dual-arm cooperate to grasp targets of different depths and different poses. The method is implemented as follows (see Fig. 1): the robot obtains the image of the workspace (with known internal parameters) through the left arm camera. The depth information of the target is first obtained through the generated mapping model. Then the target is located based on the depth information and pixel coordinate information to obtain the 3D positional parameters of the target. Finally, based on the main-auxiliary control, control commands are sent to both robot arms to realize collaborative grasp by dual-arm.

3.1 Mapping-Based Depth Acquisition

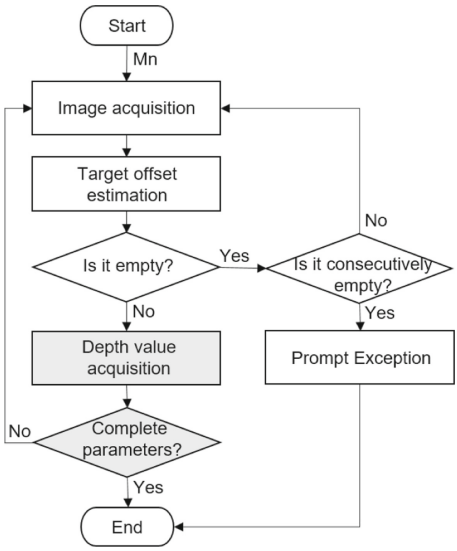

Fig. 2. Mapping-based depth acquisition flowchart.

In this paper, we use the built-in RGB camera of the robot. Figure 2 is the flow chart of this module. Among them, Mn represents the total number of markers. Get the camera image of the robot's left arm through the Robot Operating System. The target offset estimation is to obtain the offset vector of the marker in the camera coordinate system through marker detection. Then, the depth information is obtained through mapping processing, and the pixel coordinate information of each marker is saved. If the mark detection result is empty continuously, an exception needs to be prompted. The implementation method of depth value acquisition and parameter integrity judgment is described below.

Depth Value Acquisition. Obtaining accurate depth information is mission-critical. This paper uses the offset of the marker with respect to the Z axis of the left-arm camera coordinate system as approximate depth information. And the steps of first reducing volatility of depth information and then increasing accuracy are used to obtain more accurate depth values.

In order to obtain a more stable depth value, the mean method is used to reduce the volatility, and the sample mean S_a is calculated as shown in Eq. 1.

$$S_a = (\sum_{m=0}^{n} val_n)/(n+1) \tag{1}$$

where m is the counter, n is the set threshold, and val is the Z-axis offset of the marker in the camera coordinate system. When a marker information is detected in a frame of image, the value of the counter m is incremented by one, and the val is accumulated and stored in the sum. If a frame of pictures is invalid, the values of m and sum will not be changed. When the counter reaches the threshold, record S_a as the sample value, then set both m and sum to zero.

In order to obtain a more accurate depth value, the real depth is changed according to a certain rule, and a sample value corresponding to the depth value is collected. There is a certain difference between the sample value and the real value. Since the fluctuation of the sample value is reduced by Eq. 1 processing, the sample value under the same depth value is relatively stable. The corresponding relationship model is obtained by curve fitting the collected multiple sets of sample values and the real depth values. $f_d(S_a)$ represents the relationship between the depth value and the sample mean.

$$f_d(S_a) = p_1 \cdot (S_a)^3 + p_2 \cdot (S_a)^2 + p_3 \cdot (S_a) + p_4 \tag{2}$$

where S_a denotes the sample mean and p_1, p_2, p_3, p_4 are the coefficients. The mean value S_a is re-acquired as an independent variable, and the depth information d is obtained through relational model processing.

In the subsequent calculation of the 3D pose of the target, depth information and pixel information need to be used. The pixel information of the complete grab point needs to be saved here. As in the marker detection, there is a failure to detect all markers. In order to ensure the completeness of parameters, we create a parameter judger (P_J) to solve the problem of incomplete parameters, and the expression is as follows.

$$P_J = \begin{cases} 1, \forall m_i \neq 0 \\ 0, otherwise \end{cases} \tag{3}$$

where m_i represents the element in the parameter vector M, and the parameter vector M is input into the parameter judger. When P_J returns 1, it means that the parameters are complete, and when it returns 0, it means that the parameters are missing, and the image needs to be re-acquired, and the detection is performed until the complete parameters are obtained.

3.2 Depth-Based Object Positioning

Object positioning includes target location positioning and target pose estimation. The target location positioning is to obtain the position of the object in the robot world coordinate system, and the target pose estimation is to obtain the rotation vector from the robot's world coordinate system to the object's own coordinate system. Details are described below.

Target Location Positioning. The depth information obtained by the above method is combined with pixel coordinates to calculate the three-dimensional coordinate position of the target in the camera coordinate system, and then converted into coordinates in the robot world coordinate system. Take the marked center point as the target position P. Obtain the pixel coordinates $P^{'}$ of the center point of the marker through the pixel coordinates of the corner points of the marker.

In the image coordinate system, $c^{'}$ is the "center" point of the image corresponding to the optical axis, and the length of the coordinate difference between $p^{'}$ and $c^{'}$ as Eq. 4.

$$L(\Delta i)|_{i=u,v} = \Delta i \cdot dx \tag{4}$$

where u, v represent the image coordinate axes, Δi represents the coordinate difference between $p^{'}$ and $c^{'}$, and dx represents the width of the square pixel.

In the camera coordinate system (three-dimensional coordinates), based on the principle of optical imaging, according to the correspondence between the left-arm camera coordinate axis and the image coordinate axis, the target is in the camera coordinate system, about the coordinates of the X and Y axes The values are $X_{w(u)}, X_{w(v)}$.

$$X_w(i)|_{i=u,v} = d \cdot L(\Delta i)/f \tag{5}$$

where d represents the depth of the target from the camera (Eq. 2), f represents the focal length of the camera.

By calibrating the camera, the position of the target under the camera coordinate system can be derived as C_p based on the camera internal reference.

$$C_p = [X_w(u), X_w(v), d]^T \tag{6}$$

where $X_w(u), X_w(v)$ see Eq. 5, d denotes the depth information see Eq. 2.

According to the value of f_x in the camera internal reference, the value of $L(\Delta i)/f$ near the center of the image is less than 1. Therefore, the influence of the depth value error on the coordinate value will be reduced, the coordinates of the target in the camera coordinate system about the X-axis and Y-axis will be more accurate. Then convert the target position to the robot world coordinate system, and the position of the target in the world coordinate system is represented as P_w.

$$P_w = P_0 + R_{cw} \cdot C_p + R_{gw} \cdot G_c \tag{7}$$

where P_0 denotes the starting position of the left-arm planar gripper end under the robot world coordinate system. R_{cw} denotes the rotation matrix of the conversion from the camera coordinate system to the robot world coordinate system. C_p denotes the position of the target in the camera initial position coordinate system. R_{gw} denotes the rotation matrix of the conversion from the left arm planar gripper coordinate system to the robot world coordinate system. G_c denotes the position of the left-arm camera in the left-arm plane gripper coordinate system. Since the positions of the left-arm camera and the left-arm plane gripper are relatively fixed, G_c is a constant value.

Target Pose Estimation. Figure 3 shows the rotational change of the marker in the image. The upper left corner indicates the origin of the image coordinate system, the horizontal direction indicates the u-axis, and the vertical direction indicates the v-axis. The second column shows the cases when the marker is not deflected and deflected by 180°, and the pose of the marker when $v_i = v_j$ is set as the initial pose. The first column indicates that the marker is rotated counterclockwise with respect to the initial pose, and the third column indicates that the marker is rotated clockwise with respect to the initial pose. The angle marked by the red line is the angle of marker deflection.

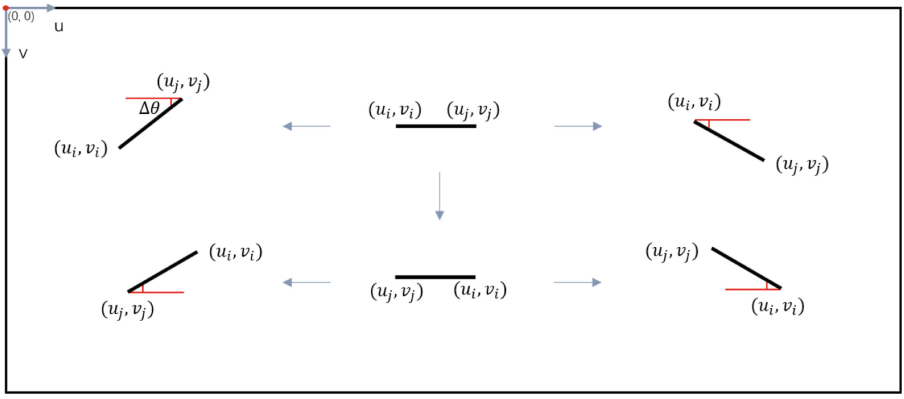

Fig. 3. Different rotation states of markers.

Since the positions of the corner points of the marker are relatively fixed, the deflection angle $\Delta\theta$ of the marker with respect to the initial pose is calculated using the fixed corner points.

$$\Delta\theta = \arctan \frac{|\ \Delta v_{ij}\ |}{|\ \Delta u_{ij}\ |} \tag{8}$$

where Δv_{ij} denotes the change of longitudinal coordinates of corner point i and corner point j, Δu_{ij} denotes the change of transverse coordinates of corner point i and corner point j, and $arctan$ represents the inverse tangent function.

When the marker is deflected counterclockwise, the slope of the line where corner point i and corner point j are located is negative in the image coordinate system. On the contrary, when the marker is deflected clockwise with respect to the initial pose, the slope of the line where corner point i and corner point j are located is positive. The direction of rotation is expressed as f_{dor}.

$$f_{dor} = \begin{cases} 1, if \Delta u_{ij} \cdot \Delta v_{ij} < 0 \\ -1, otherwise \end{cases} \tag{9}$$

where Δv_{ij} indicates the difference between the longitudinal coordinates of corner point i and corner point j, and Δu_{ij} indicates the difference between the lateral coordinates of corner point i and corner point j. A f_{dor} of 1 indicates that the marker is deflected counterclockwise with respect to the initial pose, and a f_{dor} of -1 indicates that the marker is deflected clockwise with respect to the initial pose.

In the camera coordinate system, when the marker is in the initial pose, the relative states of the camera and the marker are called the initial state. The rotation vector of the left-arm gripper coordinate system in the world coordinate system is denoted as θ_w. When controlling the robot, the θ_w parameter Euler angles need to be converted into Quaternion to control the plane gripper state.

$$\theta_w = \theta_g + f_{dor} \cdot \theta_o \tag{10}$$

where θ_g represents the rotation vector from the robot world coordinate system to the left arm gripper coordinate system when the camera is in the starting position, and θ_o represents the rotation vector when the camera is transformed from the initial position to the initial state, $\theta_o = [0, 0, \Delta\theta]^T$.

3.3 Main-Auxiliary Control Based Dual-Arm Grasping

The flow chart of dual-arm grasping based on main-auxiliary control is shown in Fig. 4. The selected target is determined based on the number of targets N. In case of multiple objects, marker pairing is required (markers are placed in ascending order, and two are selected in sequence as two grasp points for the same object). The human hand detector will output the pixel center coordinates of the hand. If the object selection condition is satisfied in three consecutive frames, the object is determined. Based on the pixel coordinates of the two grasping points of the object, it is judged whether the positional parameters need to be exchanged, and then the two grasping point position parameters of the object are assigned to the left and right robotic arms. The object selection conditions and the dual-arm cooperative control method are described as follows.

Object Selection. N is the number of the target object T. The target is determined according to whether there are multiple target objects. If N is greater

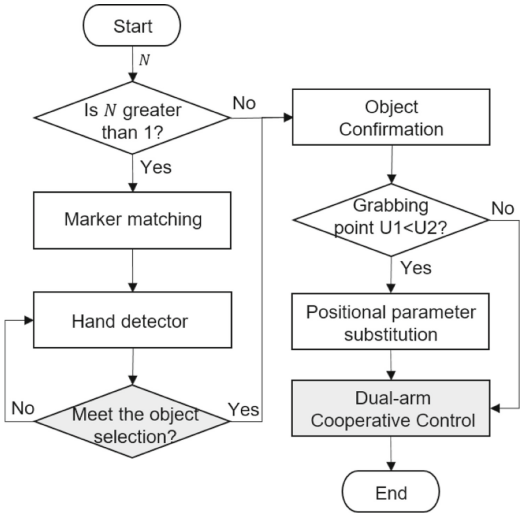

Fig. 4. Main-auxiliary control based dual-arm grasping.

than 1, since each object T_n occupies a pair of marks, where $n \in [1, N]$. The coordinates of the two marked center pixels of the object T_n are (u_{n_1}, v_{n_1}) and (u_{n_2}, v_{n_2}). In the image coordinate system, the pixel coordinate value of the right arm of the robot is smaller than the value of the left arm. And according to the rules for storing the coordinate parameters, it is necessary to compare the pixel coordinates of the two grasping points to decide whether the pose parameters should be swapped.

$$T_{grasp}(N) = \begin{cases} T_1, if N = 1 \\ T_a, otherwise \end{cases} \tag{11}$$

where $a = \{n|h(u_h, v_h) \in [u_{n_1} : u_{n_2}, v_{n_1} : v_{n_2}]\}$,$[u_{n_1} : u_{n_2}, v_{n_1} : v_{n_2}]$ represents a rectangular pixel frame consisting of (u_{n_1}, v_{n_1}) and (u_{n_2}, v_{n_2}) as diagonal points. $h(u_h, v_h)$ represents the pixel center point of the hand.

Dual-Arm Cooperative Control. The timing schematic for main-auxiliary control is shown in Fig. 5. In this paper, the main-auxiliary control (unified control clock) is used to achieve synchronization of dual-arm collaborative operation. The method uses main-auxiliary control for the control of the left and right robotic arms. The main control enables the control of both robotic arms, while the auxiliary control is used to control the right robotic arm only. The main control and auxiliary control are used simultaneously only when the cooperative operation is performed. A communication connection is used between the main and the auxiliary control. The main control transmits the joint angle of the right arm to the auxiliary control, and the auxiliary control returns the identification information after the right arm task is completed. For cooperative operation, it is necessary to unify the main and the auxiliary control clocks, which is done

by a delayed waiting strategy. Theoretically, the delay waiting time of the main control should be equal to the delay waiting time of the auxiliary control plus the propagation delay. In practice, the propagation delay is so small as to be negligible when the main control transmits information to the auxiliary control. Therefore, the main and the auxiliary control unify the control clocks by waiting for the same amount of time. The main and the auxiliary control then sends control commands to the left and right robotic arms simultaneously to ensure the synergy of the two arms.

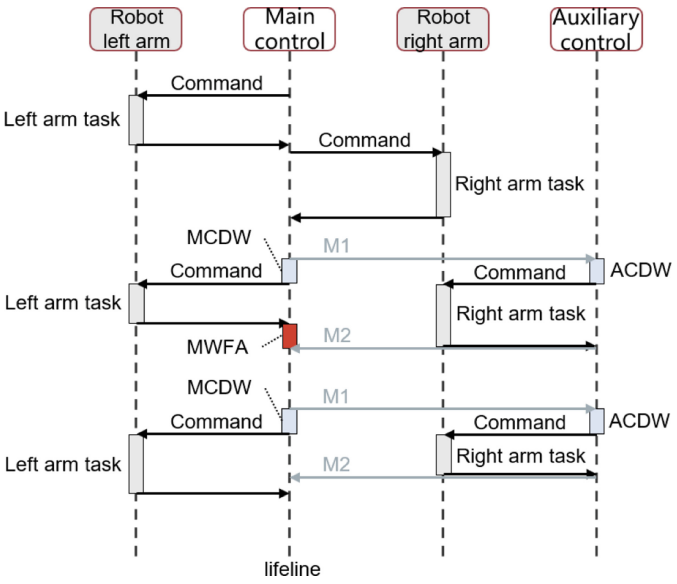

Fig. 5. The timing schematic for main-auxiliary control. ACDW: Auxiliary control delay waiting. MCDW: Main control delay waiting. MWFA: The main control waits for the auxiliary control. M1: Right arm joint angles. M2: Identifying characters.

4 Results

To verify the effectiveness of our proposed method, we conducted multiple sets of experiments with the Baxter robot as an example. The experimental configuration will be described in Subsect. 4.1. In Subsect. 4.2 we describe the experimental design. In Subsect. 4.3, we will present the experimental results and analyze the results.

4.1 Experimental Configuration

Throughout the experiments, a computer, an experimental workbench with variable height, target objects with locally graspable features and markers [4,14,18] of known width were used. And a Baxter robot with a parallel gripper equipped with an RGB camera (1280 × 800 pixels) on its left arm, as shown in Fig. 6.

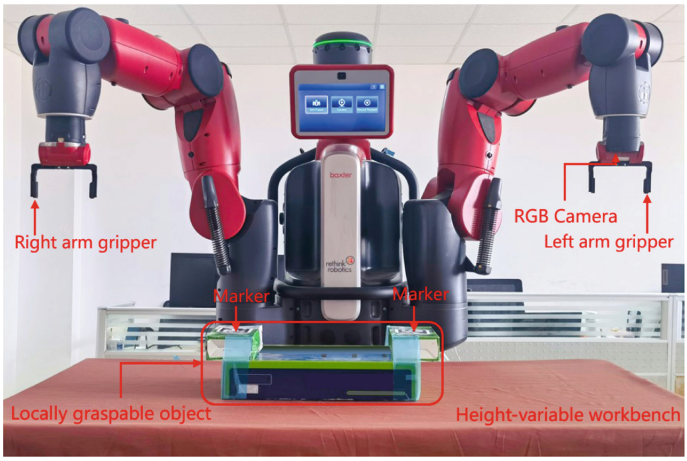

Fig. 6. Experimental scene diagram.

4.2 Experimental Design

In order to quantitatively evaluate the influence of different depths, different poses, and different environments on the dual-arm cooperative grasping method, we verify the effectiveness and interference resistance of the method in two parts. The first part is to verify the effectiveness of the method under different depths, different lighting conditions, and different positional conditions. The second part is to verify the robustness of the method under the conditions of redundant marker interference and multiple target selection.

Experiment 1 design: (1) To verify whether the method can grasp targets of different depths by collaborating with dual-arm, we set the height of the workbench to five different heights (constrained by the working space of the robot arms, the height of the worktable surface from the camera is controlled within 30–70 cm, and five different heights are set within this range). (2) To verify whether the method can successfully grasp the target in different poses on the workbench, the target attitude is divided into target without deflection and with deflection angle (as shown in Fig. 7(b) (c)), and the target position is changed according to the distribution position of Fig. 7(a). (3) To verify whether the method is affected by light conditions, five sets of experiments with different depth conditions were conducted under both natural light and light conditions.

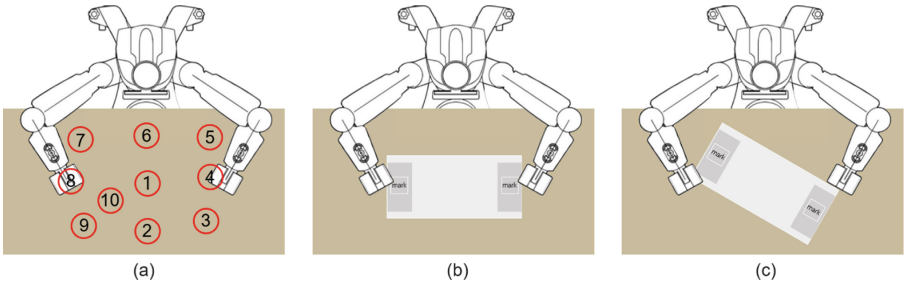

Fig. 7. Experimental target position and attitude setting schematic. (a) Schematic diagram of the approximate location distribution of the targets in the experiment. (b) Schematic diagram of the experimental target without deflection. (c) Schematic diagram of the experimental target with deflection.

Experiment 2 design: (1) Verify whether redundant markers have an effect on target grasping. A target object, three markers, and one marker as interference marker were used. (2) Verify the effectiveness of target selection. Two target objects are used with four markers, and since there are two targets, human interaction is required at this time, and the target to be grasped is determined by detecting the manually selected target objects and satisfying the target selection condition for three consecutive frames. In addition, in order to verify whether the target selection is affected by light, the experiment was conducted in a set of experiments under natural lighting and lighting conditions. The human hand detection uses Yolov5 [7] as the detection framework and weights [11].

Description: It is necessary to detect multiple frames of images when acquiring depth information (the threshold is set to 20 in this experiment). During the experiment, each set of experiments (ten grasp experiments) only acquired depth information once at the beginning, and used this depth information to complete ten grasp tasks. Since the method involves marker pairing, the following usage rules need to be defined for markers. (1) The left and right edges of the marker need to be in the same direction as the edge of the local graspable point, as shown in Fig. 8. (2) If more than two markers are used, the markers need to be ranked in ascending order by serial number, and two consecutive markers from the first one are used for the same target object.

Fig. 8. Marker usage rules.

4.3 Experimental Results and Analysis

Table 1. Experiment 1 results statistics. NL_di(i = 1, 2, . . . , 5): The height difference between the work surface and the camera under natural light conditions. EL_di(i = 1, 2, . . . , 5): The height difference between the workbench and the camera under lighting conditions. SG_n: The target is not deflected, only its position is changed, and the target grasp success rate is counted. SG_y: The target is deflected and changes its position, and the target grasp success rate statistics. RS: After placing the target, when the arms return to the initial position, sometimes due to the deviation between the distance between the dual-arm grippers and the actual width of the target, the plane gripper will take one end of the target away from the experimental desktop. RS indicates that this does not happen. NSR said that such a situation occurred. ASR: Average success rate.

Ec	SG_n	RS	SG_y	RS	ASR	
NL_d1	10/10	10/10	10/10	6/10	100%	80%
NL_d2	10/10	7/10	10/10	10/10	100%	85%
NL_d3	10/10	9/10	10/10	10/10	100%	95%
NL_d4	9/10	8/10	10/10	10/10	95%	90%
NL_d5	10/10	9/10	10/10	10/10	100%	95%
EL_d1	10/10	9/10	10/10	10/10	100%	95%
EL_d2	10/10	8/10	10/10	8/10	100%	80%
EL_d3	10/10	10/10	10/10	9/10	100%	95%
EL_d4	10/10	9/10	10/10	10/10	100%	95%
EL_d5	10/10	10/10	10/10	10/10	100%	100%
ASR	99%	89%	100%	93%	99.5%	91%

The results of Experiment 1 are shown in Table 1. Overall, the grasping success rate was 99.5% and the RS success rate was 91%. One grasping failure occurred. (1) In the experiments with different heights of the table for the grasping situation, it can be seen that (d4, d5) RS success rate is slightly higher than (d1, d2). The reason is that the target is closer to the camera under d1 condition, and the target position changes by the same distance under d1 condition, which is more obviously affected by the camera distortion. Although the experimental results of each group are slightly different, the difference is not obvious, indicating that the method is equally effective under different depth conditions. (2) The success rates of RS were 89% and 93% under the conditions of target changing position without deflection and deflection, respectively, indicating that there was no significant difference in the results when the target was in different positions, indicating that the method can be applied to targets in different positions. (3) Under the natural light and light illumination environment, the grasping success rate and RS success rate are also 89% and 93%, which indicates that the lighting effect also has no obvious effect on the method, and the method is applicable

to different lighting environments. In the case of failure to grasp during experiment 1, the right arm gripper rotates during the ascent of the dual-arm grasping target. The reason is that the optimal solution path from point A to point B is different from the optimal solution path obtained from point B to point A during the inverse kinematics solution process. When it has been adjusted by reusing the joint angle of point A, this situation does not occur anymore.

Table 2. Experiment 2 results statistics. NL_i(i = 3, 4): Use three markers under natural light conditions. EL_4: Use four markers under lighting conditions. MP: Markers pairing. TS: The correct number of times of target selection by detecting human hands. SG: The number of successful grasps of the target. RS and ASR are the same as Table 1.

Ec	MP	TS	SG	RS	ASR
NL_3	10/10	0	9/10	9/10	90%
NL_4	10/10	10/10	10/10	8/10	80%
EL_4	10/10	10/10	10/10	9/10	90%

The results of Experiment 2 are shown in Table 2. The overall task success rates were 90%, 80%, and 90%, respectively. (1) NL_3 verified the redundant marker interference situation, and the MP success rate was 100%, indicating that the redundant interference markers did not affect the results. (2) In NL_4 and EL_4 experiments, the success rates of TS, MP, and SP were 100%. This indicates that the target selection method was effective, and the illumination environment did not significantly affect the target selection. The success rates of RS in the three experiments were 90%, 80%, and 90%, respectively, which were not significantly different from the results of experiment 1. There was a grasp failure during experiment 2. The reason was that when changing the target position, the position during the move was recorded instead of the target position.

The common problems of Experiments 1 and 2, the main reasons for NRS: (1) When solving the three-dimensional coordinates, the solution is only performed once, and sometimes the solution is not accurate enough. (2) Since rubber pads are used on both sides of the flat gripper of the robotic arm, the gripper is opened, and the rubber pad will have a certain elasticity, which will also affect the placement of the target. (3) There is also a certain error in the movement of the robot arm, and each movement will be slightly deviated. Next, it is necessary to further improve the calculation accuracy of the target pose through position verification.

The method proposed in this paper was not compared with other vision-based dual-arm collaboration methods. The reason is the other methods using RGB cameras with fixed depth or using depth camera acquisition. Also due to the different experimental specific settings between us, the evaluation criteria of the results are different. Therefore, this paper is not compared with other methods.

5 Disscussion

This paper proposes a dual-arm collaboration grasp method based on RGB cameras. Firstly, the depth information model is established by RGB camera, and using the markers attached to the target to obtain the depth information of the target. Then the 3D position of the target in the robot world coordinate system is calculated. Finally, the target object is grasped by the cooperation of both arms. This method does not depend on the depth camera and can be applied to the scene with varying depth, which improves the applicability of the dual-arm collaboration method based on RGB camera. The method can be applied to depth-variant grasping tasks and autonomous assembly tasks in industrial scenes using only RGB cameras.

However, this method also has certain limitations. First, a mark needs to be attached to the local graspable point of the object, and the mark must be used according to the rules. The main reason for these limitations is that the pixel information of the local graspable points needs to be obtained through marker detection to calculate the pose. Secondly, since this method only calculates the pose change of the target on the workbench, it is not applicable when the target is inclined. Therefore, future work focuses on edge detection methods as well as target 6D grasp point estimation methods. The edge detection method is used to determine the local graspable point pixel coordinate information, and the 6D pose estimation is used to obtain the object's pose in order to expand the applicability of the method.

Acknowledgement. This work was supported by the Project of NSFC (Grant No. U1908214), Special Project of Central Government Guiding Local Science and Technology Development (Grant No. 2021JH6/10500140), the Program for Innovative Research Team in University of Liaoning Province (LT2020015), the Support Plan for Key Field Innovation Team of Dalian (2021RT06), the Science and Technology Innovation Fund of Dalian (Grant No. 2020JJ25CY001), the Support Plan for Leading Innovation Team of Dalian University (Grant No. XLJ202010) and Dalian University Scientific Research Platform Project (No. 202101YB03).

References

1. Bai, W., et al.: Dual-arm coordinated manipulation for object twisting with human intelligence. In: 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 902–908. IEEE (2021)
2. Cai, J., Cheng, H., Zhang, Z., Su, J.: Metagrasp: data efficient grasping by affordance interpreter network. In: 2019 International Conference on Robotics and Automation (ICRA), pp. 4960–4966. IEEE (2019)
3. Du, G., Wang, K., Lian, S., Zhao, K.: Vision-based robotic grasping from object localization, object pose estimation to grasp estimation for parallel grippers: a review. Artif. Intell. Rev. **54**(3), 1677–1734 (2021)
4. Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F.J., Medina-Carnicer, R.: Generation of fiducial marker dictionaries using mixed integer linear programming. Pattern Recogn. **51**, 481–491 (2016)

5. Griffin, B.A., Corso, J.J.: Learning object depth from camera motion and video object segmentation. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12352, pp. 295–312. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58571-6_18
6. Ibarguren, A., Eimontaite, I., Outón, J.L., Fletcher, S.: Dual arm co-manipulation architecture with enhanced human-robot communication for large part manipulation. Sensors **20**(21), 6151 (2020)
7. Jocher, G., et al.: ultralytics/yolov5: v5. 0-yolov5-p6 1280 models aws supervise. ly and youtube integrations. Zenodo **11** (2021)
8. Laghi, M., et al.: Shared-autonomy control for intuitive bimanual telemanipulation. In: 2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids), pp. 1–9. IEEE (2018)
9. Liang, J., Mahler, J., Laskey, M., Li, P., Goldberg, K.: Using DVRK teleoperation to facilitate deep learning of automation tasks for an industrial robot. In: 2017 13th IEEE Conference on Automation Science and Engineering (CASE), pp. 1–8. IEEE (2017)
10. Lipton, J.I., Fay, A.J., Rus, D.: Baxter's homunculus: virtual reality spaces for teleoperation in manufacturing. IEEE Robot. Autom. Lett. **3**(1), 179–186 (2017)
11. Liu, D., et al.: A novel and efficient distance detection based on monocular images for grasp and handover. In: Gao, H., Wang, X. (eds.) CollaborateCom 2021. LNICST, vol. 406, pp. 642–658. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-92635-9_37
12. Lundell, J., Verdoja, F., Kyrki, V.: Beyond top-grasps through scene completion. In: 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 545–551. IEEE (2020)
13. Medjram, S., Brethe, J.F., Benali, K.: Markerless vision-based one cardboard box grasping using dual arm robot. Multimedia Tools Appl. **79**(31), 22617–22633 (2020)
14. Muñoz-Salinas, R., Marín-Jimenez, M.J., Yeguas-Bolivar, E., Medina-Carnicer, R.: Mapping and localization from planar markers. Pattern Recogn. **73**, 158–171 (2018)
15. Ott, C., Nakamura, Y.: Employing wave variables for coordinated control of robots with distributed control architecture. In: 2008 IEEE International Conference on Robotics and Automation, pp. 575–582. IEEE (2008)
16. Punlum, V., Srisertpol, J., Khaengkam, S.: The application of double arms scara robot for deburring of PCB support plate. In: 2017 International Conference on Circuits, Devices and Systems (ICCDS), pp. 1–5. IEEE (2017)
17. Rastegarpanah, A., Marturi, N., Stolkin, R.: Autonomous vision-guided bi-manual grasping and manipulation. In: 2017 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO), pp. 1–7. IEEE (2017)
18. Romero-Ramirez, F.J., Muñoz-Salinas, R., Medina-Carnicer, R.: Speeded up detection of squared fiducial markers. Image Vis. Comput. **76**, 38–47 (2018)
19. SepúLveda, D., Fernández, R., Navas, E., Armada, M., González-De-Santos, P.: Robotic aubergine harvesting using dual-arm manipulation. IEEE Access **8**, 121889–121904 (2020)
20. Silvério, J., Clivaz, G., Calinon, S.: A laser-based dual-arm system for precise control of collaborative robots. In: 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 9183–9189. IEEE (2021)
21. Smith, C., et al.: Dual arm manipulation-a survey. Robot. Auton. Syst. **60**(10), 1340–1353 (2012)
22. Tung, A., et al.: Learning multi-arm manipulation through collaborative teleoperation. In: 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 9212–9219. IEEE (2021)

23. Wu, Q., Li, M., Qi, X., Hu, Y., Li, B., Zhang, J.: Coordinated control of a dual-arm robot for surgical instrument sorting tasks. Robot. Auton. Syst. **112**, 1–12 (2019)
24. Yang, Y., Liu, Y., Liang, H., Lou, X., Choi, C.: Attribute-based robotic grasping with one-grasp adaptation. In: 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 6357–6363. IEEE (2021)
25. Yu, X., Zhang, S., Sun, L., Wang, Y., Xue, C., Li, B.: Cooperative control of dual-arm robots in different human-robot collaborative tasks. Assem. Autom. **40**(1), 95–104 (2019)
26. Zahavi, A., Haeri, S.N., Liyanage, D.C., Tamre, M.: A dual-arm robot for collaborative vision-based object classification. In: 2020 17th Biennial Baltic Electronics Conference (BEC), pp. 1–5. IEEE (2020)
27. Zhong, F., Wang, Y., Wang, Z., Liu, Y.H.: Dual-arm robotic needle insertion with active tissue deformation for autonomous suturing. IEEE Robot. Autom. Lett. **4**(3), 2669–2676 (2019)