

## 基于 RGB-D 图像的机械臂视觉抓取算法及系统

曲睿<sup>1</sup>, 李勇<sup>1,2\*</sup>, 双丰<sup>1</sup>, 黄瀚樟<sup>1</sup><sup>1</sup>广西大学电气工程学院广西电力装备智能控制与运维重点实验室, 广西 南宁 530004;<sup>2</sup>广西大学电气工程学院广西制造系统与先进制造技术重点实验室, 广西 南宁 530004

**摘要** 机械臂可以替代人工进行抓取等工作来提高生产效率与产品质量。当前基于数据驱动算法需要大量样本训练模型, 迁移性较弱, 并要求一定的计算资源作为支撑。为解决上述问题, 建立了一个基于 RGB-D 图像的高效机械臂视觉抓取算法和系统。首先提出一种最小包围矩形 (MinBRect) 目标检测算法, 用以快速估计目标的位置信息; 然后利用 MinBRect 通过计算最小包围矩形的短轴倾角来估计抓取任务的位姿信息; 最后操纵 UR5 机械臂进行实际抓取实验。在实验中, 所提算法对 10 种目标物体的位姿估计的精度均在 85.7% 以上, 平均时间达 0.7677 s。与两种位姿估计算法进行对比, 所提算法的抓取精度和时间均有明显改善, 这表明所提算法具有较强的准确性和鲁棒性。此外, 由于无需构建数据集, 所提算法能够应用到计算资源有限、工业生产中的结构化场景, 在电力工件实验中验证了所提算法具有较好的可移植性。通过实际抓取实验进一步验证了所提机械臂系统抓取算法的有效性。

**关键词** 图像处理; 位姿估计; 机械臂抓取; 视觉定位; 目标检测

中图分类号 TP242.6

文献标志码 A

DOI: 10.3788/LOP202259.1415025

## Robotic Arm Visual Grasping Algorithm and System Based on RGB-D Images

Qu Rui<sup>1</sup>, Li Yong<sup>1,2\*</sup>, Shuang Feng<sup>1</sup>, Huang Hanzhang<sup>1</sup><sup>1</sup>Guangxi Key Laboratory of Intelligent Control and Maintenance of Power Equipment, School of Electrical Engineering, Guangxi University, Nanning 530004, Guangxi, China;<sup>2</sup>Guangxi Key Laboratory of Manufacturing System & Advanced Manufacturing Technology, School of Electrical Engineering, Guangxi University, Nanning 530004, Guangxi, China

**Abstract** To increase manufacturing efficiency and product quality, robotic arms can replace manual grabbing and other duties. Currently, data-driven algorithms need a significant number of samples to train models, which is less migratory, as well as a specific quantity of processing resources. To address the above issues, this study establishes an efficient robotic arm vision grasping algorithm and system based on RGB-D images. First, we propose a minimum bounding rectangle (MinBRect) target detection algorithm to quickly estimate the target position. Further, the MinBRect is used to calculate the minor axis inclination of the rectangle to estimate the pose for the grasping task, and finally, the UR5 robotic arm is manipulated to perform the actual grasping experiment. In the experiments, the accuracy of the positional estimation of the proposed algorithm for all 10 target objects is above 85.7%, and the average time reaches 0.7677 s. The grasping accuracy and speed of the proposed approach are greatly improved when compared to the two location estimation techniques, indicating that the proposed algorithm has high accuracy and resilience. Furthermore, because no dataset is required, the suggested technique may be used in structured situations with limited computational resources and industrial production, and it has been demonstrated to be portable in power equipment testing. Actual grasping tests are used to validate the effectiveness of the suggested grasping algorithm for the robotic arm system.

**Key words** image processing; pose estimation; robotic arm grasping; visual positioning; target detection

## 1 引言

近年来, 机器人凭借低成本、自动化的优势, 替代人力完成重复性和危险性的操作, 在保障人员安全的同时极大提升了工作效率。因此通过视觉传感器执行

抓取任务是智能机器人领域的热点研究内容之一<sup>[1]</sup>。

机器人完成抓取的首要任务是目标检测, VoxNet<sup>[2]</sup>、PointNet<sup>[3]</sup>、PointRCNN<sup>[4]</sup>等采用卷积神经网络训练大量的数据集, 由于缺乏特征的全局信息, 这类方法直接处理复杂的点云数据时占用大量的计算资源

收稿日期: 2022-03-21; 修回日期: 2022-04-25; 录用日期: 2022-05-23

基金项目: 国家自然科学基金(61720106009, 61773359)、广西制造系统与先进制造技术重点实验室开放基金(20-065-40S005)

通信作者: \*yongli@gxu.edu.cn

源。徐冠宇等<sup>[5]</sup>基于点对特征进行算法改进并将改进算法应用在机械臂抓取上。三维检测算法存在着耗时较大的问题,相比之下,二维检测算法更满足快速性的需求。张磊等<sup>[6]</sup>基于直接线性变换和非线性优化来提高位姿估计精度。Faster R-CNN 检测器<sup>[7]</sup>和 FPN<sup>[8]</sup>解决之前的部分瓶颈,但仍存在特征冗余的问题,王凤随等<sup>[9]</sup>和顾桂梅等<sup>[10]</sup>提出改进的 Faster R-CNN 来提高检测精度。YOLO 系列<sup>[11-13]</sup>将目标边框检测的问题转换为回归问题进行处理,但是该类算法存在定位精度较低和泛化能力较弱的问题。崔海华等<sup>[14]</sup>在 YOLO 基础上提出了一种利用多视结构光优化位姿估计的算法。整体来说,虽然当前目标检测方法努力使速度和精度都达到一定水平,但仍存在以下问题:一方面基于深度学习的方法需要构建高质量的数据集,但构建的位姿估计数据集较为复杂,同时网络架构复杂,需对数据样本进行迭代训练,算法耗时较长;另一方面,由于缺少深度信息,仅基于 RGB 图像进行空间内目标定位的检测算法难以满足抓取任务中对精度高的要求。

位姿估计是机器人抓取任务的又一个关键步骤,现阶段已有许多成熟方法,如基于模板的 Linemod 算法<sup>[15]</sup>等,这些方法易于实现,但需要人工设计特征提取算子,且缺乏一定的普适性。Sundermeyer 等<sup>[16]</sup>训练并提取 RGB 图像中的 6D 隐变量,对该变量与数据库中的位姿进行相似度比对,以实现物体位姿估计。PVN3D<sup>[17]</sup>、PPR-Net 系列<sup>[18-19]</sup>等方法通过复杂的网络设计提升了位姿估计精度,但也需要大量时间训练模

型。这些方法需要构建庞大的数据集,并且要已知目标的 3D 模型的关键点计算 2D 与 3D 的映射关系。

因此,针对上述问题,本文将 RGB 图像与深度信息相融合,在保证检测速度的同时,解决从二维图像映射到三维空间内定位的精度问题,并构建最小包围矩形 (MinBRect) 算法快速检测目标的姿态信息,实现对未知目标姿态信息的泛化估计。本文提出了一种基于 RGB-D 图像的高效目标检测算法 MinBRect。该算法通过检测目标最小包围矩形,快速有效地检测目标位置和姿态信息,无需高性能服务器训练,因此 MinBRect 算法在工业场景下的低成本处理器上具有一定的应用价值。本文构建了一个基于 RGB-D 图像快速检测的机械臂抓取系统。构建的系统采用眼在手上的形式,让相机能够随着机械臂移动,使得相机实时检测目标时的视野更大的同时减少遮挡,执行抓取任务时更加稳定高效。通过 10 组不同类型目标的组合抓取实验,结果表明,构建的视觉抓取系统识别速度快,抓取成功率高,具有实时性、鲁棒性和稳定性的特点。构建的系统的物体抓取实验视频链接为 <https://graymoor.github.io/MinBRect>。

## 2 视觉抓取系统设计

提出一种基于 RGB-D 图像的机械臂高效视觉抓取系统,并采取眼在手上的相机随动的方式,避免了相机固定导致存在视角遮挡的问题。系统的流程如图 1 所示。所提 MinBRect 快速检测算法检测物体的连通

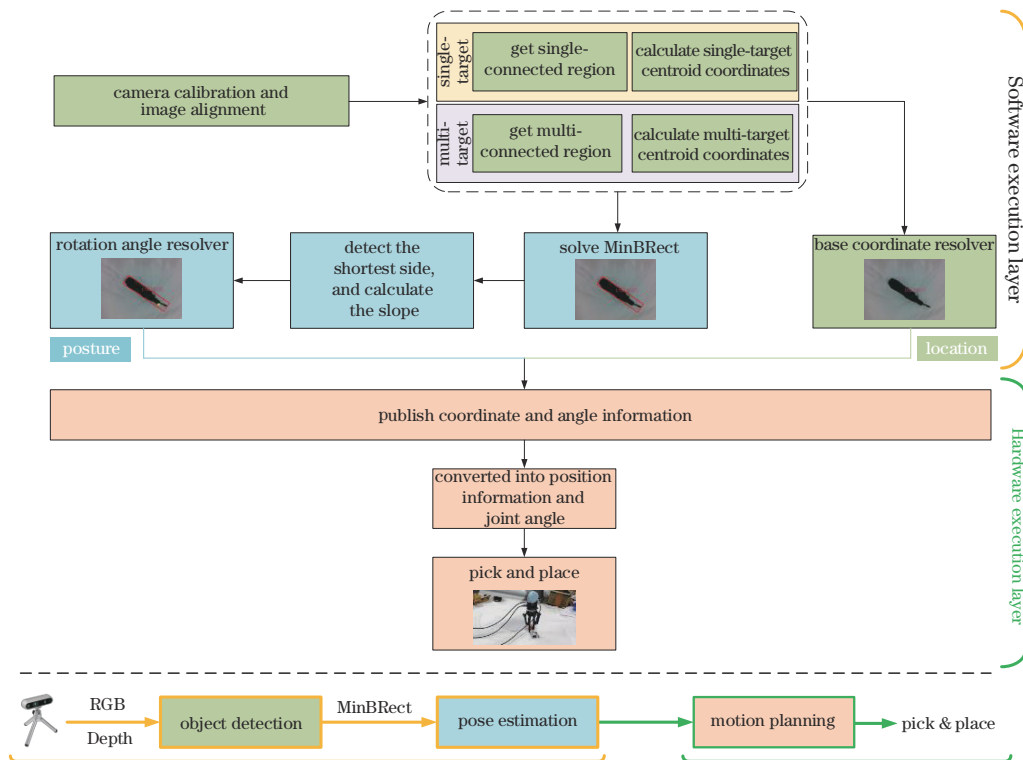


图 1 视觉抓取系统结构

Fig. 1 Structure of visual grasping system

域并进行目标的动态识别;位姿估计阶段通过计算最小包围矩形的旋转角度以调整抓取姿态;最后基于逆运动学完成机械臂的运动规划。基于数据驱动的算法需要在高性能 GPU 上训练模型,由于所提算法无需构建数据集和进行训练等工作,相较之下减少了前期准备工作的时间成本和算力成本,在小样本及未知目标的检测上具有一定优势,同时便于在较低性能的处理器的部署运行。此外,为避免传统检测方法缺少深度信息的问题,融合了 RGB 图像与深度信息。

### 2.1 硬件平台

如图 2 所示,本系统使用六自由度的 UR5 机械臂,有效负载约 5 kg,工作半径为 850 mm,适合执行如抓取等小重量任务。末端执行器采用的是 Robotiq 2f-140 二指夹爪,其具有 140 mm 的行程。视觉模块采用的是 RealSense D435i 摄像头,内置 IMU 单元 BMI055,可实现 IMU 数据和深度的实时同步,为机器人系统提供了位置感知功能。

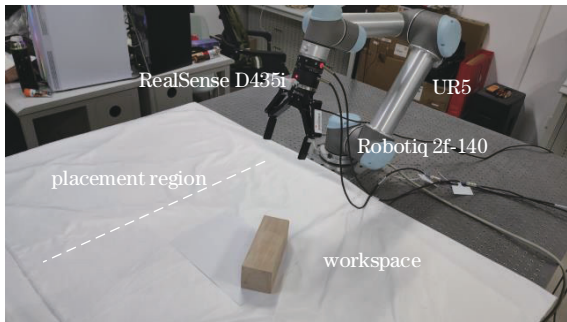


图 2 硬件平台

Fig. 2 Hardware platform

### 2.2 软件平台

ROS(robot operating system)是一个用于设计或开发机器人的开源的系统。在 ROS 中各节点通过话题发布和接收进行通信,本系统利用 ROS 与机械臂通信并执行运动规划。OpenCV 是一个跨平台的机器视觉库,因轻量化的特点和丰富的功能被广泛应用,同时具有高可移植性,也能快速应用在不同的平台上。本系统利用 OpenCV 的特点实现快速检测。

## 3 基于 MinBRect 的视觉抓取系统

基于深度学习的目标检测依赖数据集,数据集中局限的目标也限制算法的泛化能力。为克服上述问题,通过改进的大津法(OTSU)和多采样连通域的目标检测算法快速获取位置信息,同时为了准确描述目

标的姿态,提出 MinBRect 算法进行位姿估计。

### 3.1 基于连接域的目标检测

#### 3.1.1 改进的 OTSU 和连通域的快速目标检测

##### 1) 基于边缘特征信息改进的 OTSU

采取的目标检测流程如图 3 所示。为了在无需目标 3D 模型的情况下准确获取目标的距离信息,首先对齐视频流的 RGB 图线程与深度图线程,同时为了减少光照、抖动等影响,算法对图像进行了预处理<sup>[20-21]</sup>。基于 Retinex 理论进行图像预处理,先将 RGB 图像转换至 HSI 彩色空间;然后利用多尺度高斯滤波函数与图像的亮度通道 I 进行卷积,提取图像中的光照分量;再根据光照分量的分布特性动态调整校正系数,实现自适应 Gamma 矫正,抑制干扰的影响;基于改进的 OTSU 方法计算类间差异最大时的预期阈值,对于每个可能的阈值,计算像素的最大类间方差用于分割目标和背景,将 Sobel 检测算子提取的边缘信息作为确定阈值的约束条件。利用 Sobel 检测算子提取边缘点  $I(x, y)$ ,建立二维直方图,然后使用截距阈值分割,最后求得类间方差最大时的阈值。约束边缘点的计算公式为

$$(x_k, y_k) = \arg \max \left( \left| \frac{Ax + By + C}{\sqrt{A^2 + B^2}} \right| \right), k = 1, 2, (1)$$

式中:  $A=1, B=-1, C=0$ ;  $k$  代表上层和下层两个区域的点。任何一个像素都可以分为两类:

$$(x_i, y_i) \in \begin{cases} D_0, & x_k + y_k \leq T \\ D_1, & x_k + y_k > T \end{cases}, (2)$$

式中:  $x_i$  和  $y_i$  分别代表 Sobel 边缘检测的水平图像和垂直图像的灰度值;  $D_0$  为目标区域;  $D_1$  为背景区域;  $T$  为截距阈值。则目标类和背景类的概率分别为  $w_0(T) = \sum_0^T p(T_1)$  和  $w_1(T) = \sum_T^{l-1} p(T_1)$ ,  $l$  是灰度级。

此时,类内灰度均值  $u_0$  和  $u_1$ 、总灰度均值  $\mu$  分别为

$$\begin{cases} u_0 = \sum_0^T T_1 \cdot \frac{p(T_1)}{w_0} \\ u_1 = \sum_T^{l-1} T_1 \cdot \frac{p(T_1)}{w_1} \\ \mu = \sum_0^{l-1} T_1 \cdot p(T_1) \end{cases}, (3)$$

使类间方差  $L_1$  最大时的  $T$  即为所求:

$$L_1 = \max_{0 \leq T \leq l-1} \left[ w_0(u_0 - \mu)^2 + w_1(u_1 - \mu)^2 \right]. (4)$$

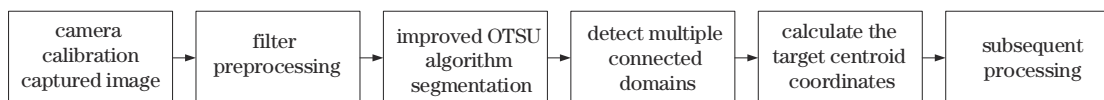


图 3 目标检测过程

Fig. 3 Target detection process

2) 基于多采样约束的连通域标注

在提取二值图像中的特定信息后,所提算法检测图像信息中目标的连通域,从而可以高效地统计图像中被标记的特定区域的面积。在连通域中求得相机坐标系下目标的质心坐标  $(X_c, Y_c)$ ,结合从深度图获得的相机坐标系下对应于连通域的  $Z_c$ ,再转换为基坐标系后进行抓取。标记后的质心点如图 4 所示。

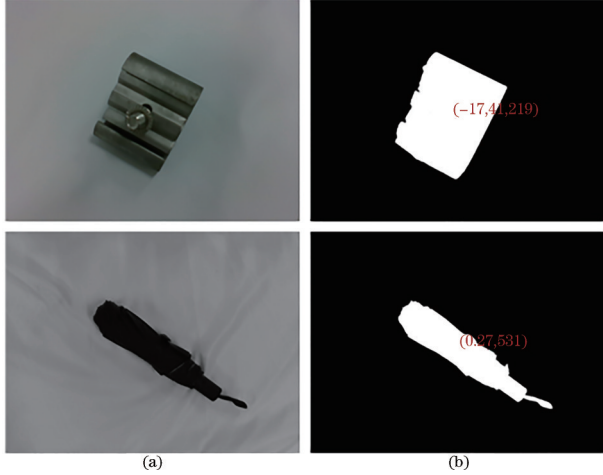


图 4 质心点标记。(a)原始图像;(b)目标质心点坐标标记  
Fig. 4 Marking of centroid points. (a) Original images;  
(b) marking of the target centroid coordinates

若图像中共有  $m$  个连通域,统计图像中的连通域  $S_m$  并用标记进行区分,具体算法流程如图 5 所示。

```

Input: scan Fig. 4 processed image pixels to find white points with pixel
 $p(x, y) = 1$ :
while not get all connected domains  $S_m$  do:
a. the seed pixel  $p(x, y)$  gives an initial label, and pushes all foreground pixels
adjacent to the seed onto the stack;
b. the pixel at the top of the popup stack assigns the same label, then push all the
pixels adjacent to the pixel into the stack;
c. get all the pixel values in the connected domain  $S_i$ 
End
while not count the  $p_i(x, y_i)$  of all pixels in the connected domain do:
calculate the centroid coordinates:  $c_x = \frac{\sum m_{i0}x_i}{\sum m_{i0}}, c_y = \frac{\sum m_{i0}y_i}{\sum m_{i0}}$ 
End
Output: pixel coordinates are converted to camera coordinates:

$$X_c = \frac{(x - c_x)Z_c}{f_x}, Y_c = \frac{(y - c_y)Z_c}{f_y}$$


```

图 5 连通域标记算法流程

Fig.5 Flowchart of connected domain marking algorithm

为使连通域标记算法更好地适用特定场景,引入多采样约束函数。要得到最佳中心点坐标,则使多采样约束函数  $L_2$  达到最小:

$$L_2 = \min \left\{ \frac{1}{m} \left[ \sum_0^{S_m} (c_x - x_i)^2 + \sum_0^{S_m} (c_y - y_i)^2 \right] \right\}. \quad (5)$$

3.1.2 基坐标求解

为准确控制机械臂到达目标位置,需进行坐标转换。要想获得工作场景中任意点在相机坐标系下的三维坐标,需要将相机坐标系转换到机械臂的基坐标系下,转换关系为

$$P_C = [R|t] P_B = H_{\text{end-off}} M_{\text{base-off}}^{\text{end-off}} P_B, \quad (6)$$

式中:  $P_C$  代表相机坐标系下的坐标;  $P_B$  代表机械臂的基坐标系下的坐标;  $[R|t]$  代表旋转平移矩阵;  $H_{\text{end-off}}$  为相机外参;  $M_{\text{base-off}}^{\text{end-off}}$  为相机坐标系到机械臂的基坐标系的转移矩阵。

设任意点在相机坐标系下表示为  $P_C = [X_C \ Y_C \ Z_C \ 1]$ , 该点在基坐标系下表示为  $P_B = [X_B \ Y_B \ Z_B \ 1]$ , 则转换关系为

$$\begin{pmatrix} X_B \\ Y_B \\ Z_B \\ 1 \end{pmatrix} = [R|t]^{-1} \cdot \begin{pmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{pmatrix}. \quad (7)$$

3.2 基于 MinBRect 的位姿估计

3.2.1 计算最小包围矩形

为方便夹爪抓取,需寻找最佳的位姿角度。首先要求得目标的最小包围矩形。先获取目标边界轮廓,再获取该目标中最小外接矩形的中心及矩形顶点坐标<sup>[22]</sup>,对相应的目标进行标注,如图 6 所示。

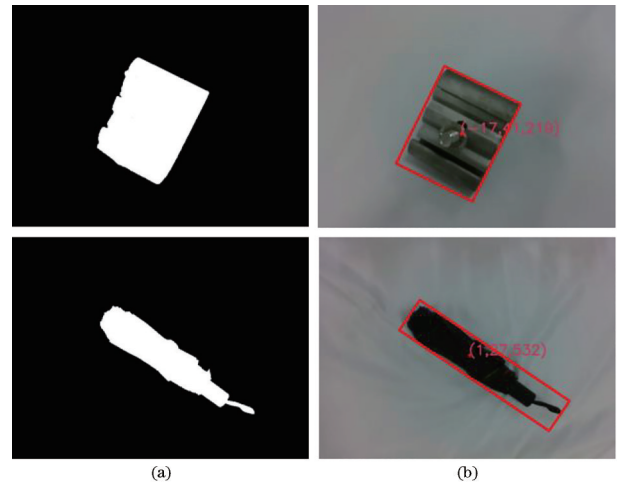


图 6 提取的最小包围矩形。(a)待检测的图像;(b)所提算法检测的目标的最小包围矩形

Fig. 6 Extracted minimum bounding rectangle. (a) Images to be detected; (b) minimum bounding rectangle detected by the proposed algorithm for targets

1) 依次扫描图像的各列  $i (i = 1, 2, \dots, N)$ , 得到每列中目标的起始坐标  $(x_{i1}, y_{i1})$  和终点坐标  $(x_{i2}, y_{i2})$ , 进一步计算该列内目标的重心坐标  $(x_i, y_i)$ :

$$x_i = \frac{x_{i1} + x_{i2}}{2}, y_i = \frac{y_{i1} + y_{i2}}{2}. \quad (8)$$

2) 把目标各列重心坐标  $(x_i, y_i)$  作为一组数据, 利用最小二乘法进行直线拟合, 得到表示水平主轴初

始位置的直线方程  $f_1(x_i) = a_1x_i + b_1$ , 同理依次扫描图像各行, 拟合垂直主轴的初始位置的直线方程  $f_2(y_i) = a_2y_i + b_2$ , 其中  $a_1, a_2, b_1, b_2$  分别代表线性拟合系数。

3) 分别求出水平和垂直方向两条平行线之间的距离, 得到外接矩形的宽度  $w_{\min\text{bbx}}$  和高度  $h_{\min\text{bbx}}$ , 计算出四个坐标点集  $C_{\text{box}} = (x_i, y_i) (i=1, 2, 3, 4)$ 。方程  $F_{x,i}^2(x_i, y_i)$  和  $F_{y,i}^2(x_i, y_i)$  分别代表中心点在 X 和 Y 方向的全局最小二乘。由此可以根据最小二乘法来优化中心点坐标:

$$\begin{cases} F_{x,i}^2(x_i, y_i) = \sum_{i=1}^N [x_i - f_2(y_i)]^2 \\ F_{y,i}^2(x_i, y_i) = \sum_{i=1}^N [y_i - f_1(x_i)]^2 \end{cases} \quad (9)$$

### 3.2.2 计算位姿角

在获得最小包围矩形后, 需要检测旋转角度以估计目标的位姿角。令位姿角  $\theta$  绕 X 轴逆时针旋转。如图 7 所示, 采用水平轴和短轴间的夹角。最小包围矩形的长轴是  $h_{\min\text{bbx}}$ , 短轴是  $w_{\min\text{bbx}}$ , 则位姿角  $\theta$  应该是水平轴和短轴之间的角度。位姿角的计算过程如下。

1) 寻找最短轴。比较长度  $r_1 = \sqrt{(C_{\text{box}}[1] \cdot x - C_{\text{box}}[0] \cdot x)^2 + (C_{\text{box}}[1] \cdot y - C_{\text{box}}[0] \cdot y)^2}$  与  $r_2 = \sqrt{(C_{\text{box}}[3] \cdot x - C_{\text{box}}[0] \cdot x)^2 + (C_{\text{box}}[3] \cdot y - C_{\text{box}}[0] \cdot y)^2}$  的大小, 取较小的边为最短轴。

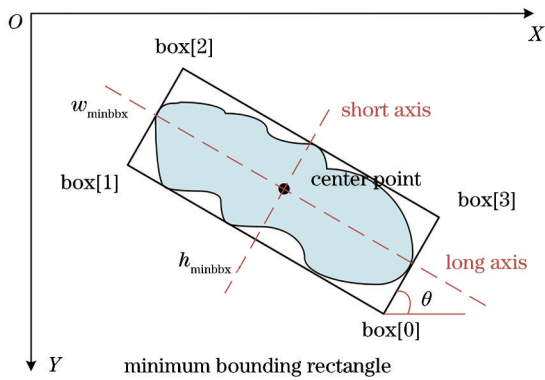


图 7 位姿角计算示意图

Fig. 7 Schematic of pose angle calculation

2) 计算旋转位姿角, 表达式为

$$\theta = \begin{cases} \arctan \frac{C_{\text{box}}[1] \cdot y - C_{\text{box}}[0] \cdot y}{C_{\text{box}}[1] \cdot x - C_{\text{box}}[0] \cdot x}, & r_1 < r_2 \\ \arctan \frac{C_{\text{box}}[3] \cdot y - C_{\text{box}}[0] \cdot y}{C_{\text{box}}[3] \cdot x - C_{\text{box}}[0] \cdot x}, & r_1 > r_2 \end{cases} \quad (10)$$

考虑到使末端夹爪旋转角度最小, 可令  $\theta \in [-90^\circ, 90^\circ]$ 。图 8 为计算得到的不同目标物的位姿角度。

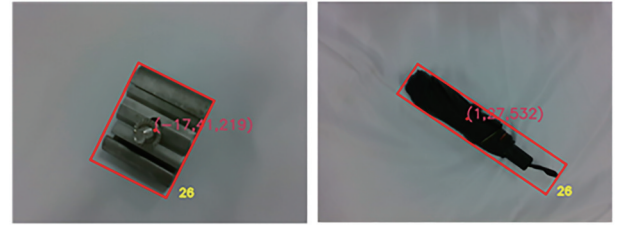


图 8 位姿角标记

Fig. 8 Marking of pose angle

为包含对位置和姿态信息的约束, MinBRect 引入约束函数  $L_3$ :

$$L_3 = \min \sum_{i=0}^N [F_{x,i}^2(x_i, y_i) + F_{y,i}^2(x_i, y_i)] + \min \theta. \quad (11)$$

由式(4)、(5)、(11), 则可以得到约束函数  $L_{\text{Total}}$ :

$$L_{\text{Total}} = L_1 + L_2 + L_3, \quad (12)$$

当三个约束条件都达到极值时, 得到 MinBRect 的最佳估计姿态信息。

### 3.3 基于逆向运动学的运动规划

实验系统采用眼在手上的机械臂抓取结构。基于 UR5 机械臂的 Denavit-Hartenberg (DH) 参数进行逆向运动学分析, 将末端执行器从初始位置移动到目标位置, 计算出末端执行器在目标点时各关节角度。UR5 在 DH 参数坐标系的结构示意图如图 9 所示, 对应的 DH 参数如表 1 所示。表 1 中,  $d_j$  和  $\theta_j$  分别表示 Z 轴方向的平移和旋转,  $a_j$  和  $\alpha_j$  表示 X 轴方向的平移和旋转,  $j$  为关节号。

根据各关节之间的几何关系计算关节角度。设  $(p_x, p_y)$  是空间坐标系中某一点的坐标, 由  $\tan \theta = p_y/p_x$  关系可以建立方程:

$$-p_x \sin \theta + p_y \cos \theta = d, \quad (13)$$

表 1 UR5 的 DH 参数

Table 1 DH parameters of UR5

Joint number $j$	Translation $d_j$ along the Z-axis	Translation $a_j$ along the X-axis	Rotation $\alpha_j$ along the X-axis	Rotation $\theta_j$ along the Z-axis
1	$d_1 = 89.459$	0	$\pi/2$	$\theta_1$
2	0	$a_2 = -425$	0	$\theta_2$
3	0	$a_3 = -392.25$	0	$\theta_3$
4	$d_4 = 109.15$	0	$\pi/2$	$\theta_4$
5	$d_5 = 94.65$	0	$\pi/2$	$\theta_5$
6	$d_6 = 82.3$	0	0	$\theta_6$

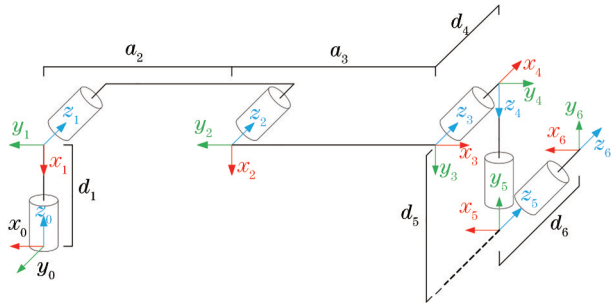


图9 UR5在DH参数坐标系的结构示意图

Fig. 9 Schematic of UR5 structure on DH coordinate system

进行三角恒等变换,令

$$\rho = \sqrt{p_x^2 + p_y^2}, \quad \vartheta = \text{atan2}(p_x, p_y), \quad (14)$$

式中:  $p_x, p_y$  分别定义为  $p_x = \rho \cos \vartheta, p_y = \rho \sin \vartheta$ ;  $\text{atan2}(\cdot)$  是 MATLAB 中的正切函数, 范围为  $[-\pi, +\pi]$ 。将上述信息代入式(13), 可以得到

$$\cos \theta \sin \vartheta - \sin \theta \cos \vartheta = d/\rho. \quad (15)$$

于是关节角可以表示为

$$\theta = \text{atan2}(p_x, p_y) - \text{atan2}(d, \pm \sqrt{p_x^2 + p_y^2 - d^2}). \quad (16)$$

## 4 实验结果与分析

### 4.1 实验环境

为验证本实验的泛化性, 采取单目标抓取和多目标抓取两种形式。选取了 10 种不同的抓取对象: 一类是规则物体, 包括木块、盒子及可乐罐; 另一类是不规则物体, 包括连接件、订书机、螺丝刀及雨伞, 和三种常见的电力工件, 包括楔形线夹、C 形线夹及单芯固定夹。

### 4.2 结果与分析

第一阶段是抓取实验中的目标检测阶段, 如图 10(a) 所示。目标的位置坐标和位姿角度的计算如图 10(e) 所示。其次控制机械臂移动到目标物上方, 如图 10(b) 所示。然后, 进入位姿调整阶段, 如图 10(c) 所示, 调整末端夹爪位姿角度准备抓取。最后当位姿调整后, 进入目标抓取阶段, 如图 10(d) 所示, 木块被抓取并放置在一个特定的区域。完成上述一组过程记录为实验成功。

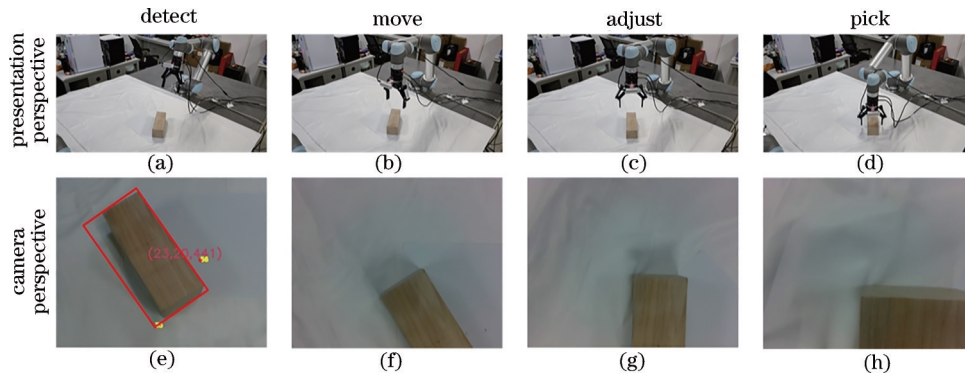


图 10 抓取过程及过程视角。(a)~(d)机械臂抓取过程;(e)~(h)对应视角

Fig. 10 Grasping process and process perspective. (a)~(d) Robotic arm grasping process; (e)~(h) corresponding perspectives

#### 4.2.1 单目标抓取实验

##### 1) 规则物体

对规则的 3 种物体(木块、盒子和可乐罐)进行每

组 7 次, 共 21 次抓取实验。表 2 展示了部分实验数据, 实验成功率为每组 7 次成功抓取的结果。其中, 实际距离指目标质心到机械臂基坐标的距离, 实际位姿角

表 2 规则物体抓取结果

Table 2 Grasping results for regular objects

Parameter	Wood block				Box				Coke can			
Actual pose angle $\theta / (^\circ)$	30		55		30		55		30		55	
Measuring pose angle $\theta' / (^\circ)$	33	35	53	57	36	38	51	57	29	27	56	57
Angle error $\Delta\theta / (^\circ)$	3	5	-2	2	6	8	-4	2	-1	-3	1	2
Actual distance $L / \text{cm}$	63	59	61.8	52.8	53.1	65.2	53.8	64.9	56.1	61.2	56.3	61.8
Measured distance $L' / \text{cm}$	59.5	56.6	65.1	51.7	53	69.9	52.3	68.3	58.1	72.3	58.2	69.2
Distance error $\Delta L / \text{cm}$	3.54	-2.4	3.3	-1.1	-0.1	4.7	-1.5	3.4	2	11.1	1.9	7.4
Success rate / %	100(7/7)				100(7/7)				85.7(6/7)			
Algorithm running time / s	0.8436	0.8325	0.7344	0.8133	0.6719	0.6903	0.9184	0.7878	0.6777	0.7995	0.7265	0.7939
Grasping quality (Excellent is E, Good is G)	E	G	E	E	G	E	E	E	E	-	G	G

指目标长轴与机械臂 X 轴的夹角,测量距离与测量位姿角是指通过上述算法获得的数值。抓取质量结果是指在抓取成功的前提下,目标质心处在夹爪中心,目标位姿角与实际基本一致记为好,否则即为良。

2) 不规则物体

对不规则的 4 种物体(连接件、订书机、螺丝刀和

雨伞)和 3 种不规则电力工件(楔形线夹 Equipment-1、C 形线夹 Equipment-2、单芯固定夹 Equipment-3)进行每组 7 次,共 49 次抓取实验。表 3 和表 4 展示了部分实验数据,实验成功率为每组 7 次成功抓取的结果。

从表 2~4 可以看出,所提算法进行视觉抓取时耗时较低,目标检测速度快,对日常物品及电力工件目标

表 3 不规则物体抓取结果  
Table 3 Grasping results for irregular objects








Parameter	Connector 				Stapler 				Screwdriver 				Umbrella 			
	30	55	30	55	30	55	30	55	30	55	30	55	30	55	30	55
Actual pose angle $\theta / (^{\circ})$																
Measuring pose angle $\theta' / (^{\circ})$	27	24	51	60	38	35	56	55	29	33	55	56	35	28	60	51
Angle error $\Delta\theta / (^{\circ})$	-3	-6	-4	5	8	5	1	0	-1	3	0	1	5	-2	5	-4
Actual distance $L / \text{cm}$	56.2	62.2	56.1	65.6	54.6	62.8	55.4	63.2	61.9	55.4	61.4	55.1	65.3	58.2	64.8	57.2
Measured distance $L' / \text{cm}$	57	66.4	58.4	69.9	51.7	65.2	56.7	68.1	61.1	53.7	63.6	52.3	59.4	60.3	67.7	61.1
Distance error $\Delta L / \text{cm}$	0.8	4.2	2.3	4.3	-2.9	2.4	1.3	4.9	-0.8	-1.7	2.2	-2.8	-5.9	2.1	2.9	3.9
Success rate / %	71.4(5/7)				71.4(5/7)				57.1(4/7)				85.7(6/7)			
Algorithm running time / s	0.7905	0.7913	0.7907	0.7886	0.7887	0.8318	0.7887	0.7881	0.7879	0.7274	0.7874	0.7881	0.7278	0.7870	0.7883	0.7301
Grasping quality (Excellent is E, Good is G)	E	-	E	G	G	G	E	-	-	E	-	E	E	E	E	E

表 4 电力工件抓取结果  
Table 4 Grasping results for electric equipments

Parameter	Equipment-1 				Equipment-2 				Equipment-3 			
	30	55	30	55	30	55	30	55	30	55	30	55
Actual pose angle $\theta / (^{\circ})$												
Measuring pose angle $\theta' / (^{\circ})$	36	34	50	53	30	28	53	54	33	28	58	57
Angle error $\Delta\theta / (^{\circ})$	6	4	-5	-2	1	-2	-2	-1	3	-2	3	2
Actual distance $L / \text{cm}$	54.8	57.2	62.7	58.4	56.8	60.2	54.4	61.7	59.6	57.3	65.2	60.3
Measured distance $L' / \text{cm}$	63.33	61.77	60.22	56	57.67	62.07	58.24	60.66	57.47	59.78	63.22	52.76
Distance error $\Delta L / \text{cm}$	8.53	4.57	-2.48	-2.4	0.87	1.87	3.84	-1.04	-2.13	2.48	-1.98	-7.54
Success rate / %	85.7(6/7)				100(7/7)				100(7/7)			
Algorithm running time / s	0.9027	0.8294	0.7518	0.8082	0.6831	0.7432	0.8384	0.6678	0.6831	0.8105	0.7457	0.8018
Grasping quality (Excellent is E, Good is G)	-	E	E	G	G	G	E	G	G	E	G	-

的平均检测时间为 0.775 s, 能够满足工业场景中对目标检测的实时性要求, 做到快速的视觉伺服检测。此外, 表 2 中所提算法对规则物体的抓取成功率在 85% 以上, 比对不规则物体的成功率略高, 这是由于抓取规则物体的容错率较高, 容易解算规则物体的质心的准确结果。从整体上来看, 所提算法对两类物体的抓取平均成功率为 81.7%。对于抓取质量, “优” 的比例较高, 所提算法对电力工件目标的平均准确率较高, 平均用时为 0.772 s, 表明所提算法在未知场景对未知物体的抓取效果方面有突出优势。图 11 是单目标抓取实验的展示。

图 12 所示的位置与角度的误差主要来自物体的阴影和鸟瞰视角, 它们改变了部分目标的最小包围矩形, 两种误差耦合性低, 没有明显的相关性, 因此相互影响小。对比表 2 和表 3 的位置误差可以看出, 对不规则物体抓取的距离误差较大些, 从图 12 展示的整体来

看误差分布比较集中, 位置误差为  $\pm 3$  cm, 角度误差为  $4^\circ$ , 相较于大多数夹爪的行程距离, 误差在行程的占比较小, 即在上述误差范围内所提算法可以满足大部分抓取场景。

4.2.2 多目标抓取实验

对比两种规则物体(盒子、可乐罐)、两种不规则物体(订书机、螺丝刀)和电力工件混合场景, 进行每组 7 次共 35 次实验, 实验结果如表 5 所示, 距离与角度的测量方式与单目标场景相同, 然后求平均误差。图 13 展示了包含电力工件组合场景的多目标抓取过程的实验。

在多目标物体抓取实验中, 从表 5 可以看出, 对规则物体抓取的成功率较不规则物体的要高一些, 对盒子和可乐罐的成功率为 85.7%, 对订书机与螺丝刀的成功率仅为 71.4%。对比二者的平均距离误差, 盒子(1.257 cm)与可乐罐(1.357 cm)的误差都在 1.5 cm



图 11 单目标抓取实验展示。(a)检测目标;(b)移动到目标;(c)抓取目标;(d)放置目标

Fig. 11 Demonstration of single-target grasping experiment. (a) Target detection; (b) moving to target pose; (c) target grasping; (d) target placement

表 5 多目标抓取结果

Table 5 Grasping results for multi objects

Parameter	Box	Cola can	Stapler	Screwdriver	Combination
Average angle error $\Delta\theta / (^\circ)$	-0.857	1.428	-2.429	-1.285	1.519
Average distance error $\Delta L / \text{cm}$	1.257	1.357	2.871	1.585	1.943
Success rate / %	85.7(6/7)	85.7(6/7)	71.4(5/7)	71.4(5/7)	85.7(6/7)
Average running time / s	1.259				



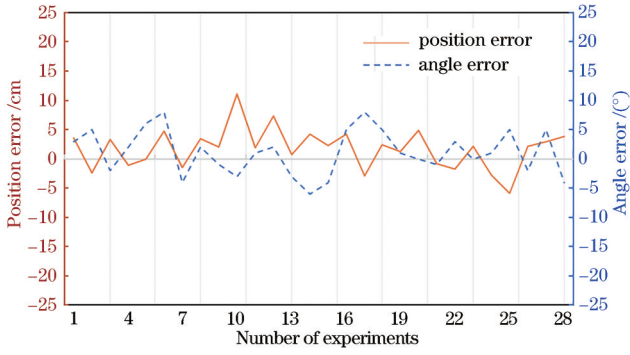


图 12 位置误差与角度误差  
Fig. 12 Position error and angle error

内,较订书机(2.871 cm)与螺丝刀(1.585 cm)的误差稍低一些;对比二者的平均角度误差,盒子(-0.857°)的误差最低,而不规则的订书机(-2.429°)最高,主要是因为规则物体的质心求解结果与实际物体的质心基本吻合,不规则的质心由于视角问题等因素产生了一定的偏差。相较于规则物体,对电力工件混合场景的成功率为 85.7%,平均角度误差与位置误差稍高一些。同时,多目标的算法耗时较单目标稍高一些,但没有出现目标加倍导致耗时加倍的现象。这也验证了所提算法在识别速度上的明显优势以及该机械臂抓取系统具有良好的鲁棒性,对未知场景下的不同物体的识别抓取有较强的泛化能力。

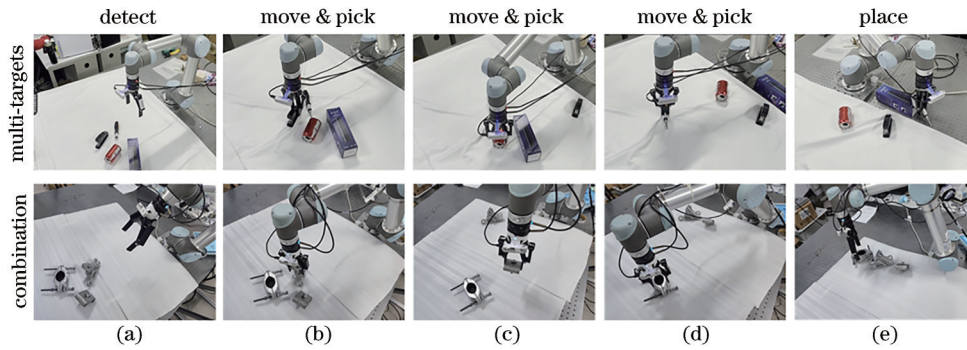


图 13 多目标抓取实验展示。(a) 检测目标;(b)~(d)移动到目标并抓取;(e)放置目标

Fig. 13 Demonstration of multi-target grasping experiment. (a) Target detection; (b)~(d) moving to target pose and grasping the targets; (e) target placement

#### 4.2.3 对比实验

使用了 ROS 中用于物体识别的工具 Object Recognition Kitchen (ORK)。ORK 包含各种识别算法,包括 Linemod 和 tabletop 等。对比 Linemod 算法、基于经典直接线性变换 (DLT) 和非线性优化的改进

位姿估计算法<sup>[6]</sup>、所提算法,这些算法都将图像坐标系转换为空间坐标系用于实际抓取实验。在实验中,测量不同位置目标的抓取角度和相对距离,在同样条件下做了 7 组对比实验。表 6 为 ORK 框架下的 Linemod 算法、文献[6]中的算法和所提算法的实验结果对比。

表 6 三种算法的对比实验结果

Table 6 Comparative experimental results of three algorithms

Algorithm	Average angle error $\Delta\theta / (^\circ)$	Average distance relative error / %	Success rate / %	Average running time / s
Linemod	—	35.6	0 (0/7)	0.7651
Algorithm in Ref. [6]	1.647	2.14	71.4 (5/7)	0.7613
Proposed algorithm	1.428	1.96	85.7 (6/7)	0.7494

由表 6 可看出:所提算法在检测速度上比 Linemod 和文献[6]中的算法略快;与所提算法的平均位置相对误差(1.96%)相比,文献[6]中的算法的平均位置相对误差为 2.14%,Linemod 的平均位置相对误差高达 35.6%,因此实际抓取的失败率很高;文献[6]中的算法的相对角度误差高所提算法 0.219°。相比之下所提算法在估计位置和姿态角度上更精准。图 14 展示了一个对比实验结果,在抓取可乐罐时,Linemod 的成功率为 0%,而所提算法为 85.7%。

表 7 为三种算法处理不同目标的对比结果,SA 表示成功率,AAS 表示平均运行时间。从表 7 来看,

Linemod 的成功率普遍低于所提算法,平均运行速度也比较慢。一方面,Linemod 对特征尺度变化很敏感,在实际的抓取实验中,Linemod 在大多数情况下难以识别目标,所预测的部分坐标不准确;另一方面,由于 Linemod 容易出现模板匹配错位,抓取效果并不好。同时可以看出,文献[6]中的算法的执行效率稍慢,所提算法在预测精度上与其相比也有一定优势。通过实验验证,所提算法在实际抓取实验中更具优势。

#### 4.2.4 消融实验

为了验证改进的 OTSU 对目标分割的有效性,对

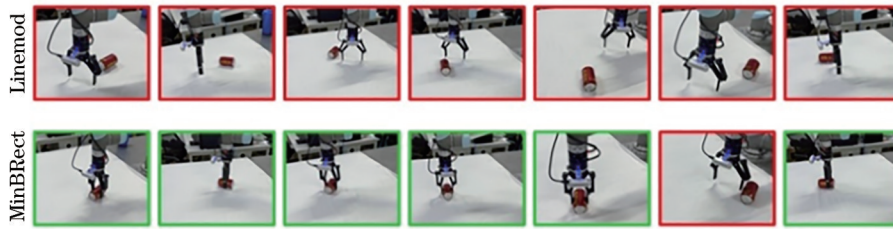


图 14 对比实验展示,红色框代表抓取失败,绿色框代表抓取成功

Fig. 14 Comparative experiment display, the red box represents grasping failure, the green box represents grasping success

表 7 三种算法处理不同目标的对比结果

Table 7 Comparison results of three algorithms for different targets

Algorithm Parameter		Wood block	Box	Coke can	Connector	Stapler	Screwdriver	Umbrella
Linemod	SA / %	14.3	28.5	0	14.3	28.5	0	14.3
	AAS / s	0.8455	0.8030	0.7651	0.8965	0.8343	0.8437	0.7846
Algorithm in Ref. [6]	SA / %	86	100	71.40	57.10	85.70	57.10	71.40
	AAS / s	0.8248	0.7782	0.7601	0.8053	0.8142	0.7819	0.7671
Proposed algorithm	SA / %	100	100	85.7	71.4	71.4	57.1	85.7
	AAS / s	0.8060	0.7671	0.7494	0.7903	0.7993	0.7727	0.7583

所提算法、K-means、OTSU 等算法进行了比较,灰度图作为输入。实验图包括单目标和多目标的 8 个场景下的 224 张测试图。根据表 8 的结果,K-means 方法的 intersection over union(IoU)优于 OTSU 方法,但在速度方面略微慢一些。对于 K-means+OTSU 方法,由于 K-means 对聚类中心敏感,所以用 OTSU 来选择初始聚类中心,然后用 K-means 聚类获得分割结果。K-means+OTSU 方法结合了 K-means 和 OTSU,在精度上有所提高,但耗时严重。所提算法在速度和精度上有明显的提高,这验证了改进的 OTSU 方法在目标分割中的有效性,并为 MinBRect 提供了快速可靠的信息。

表 8 目标分割消融实验结果

Table 8 Target segmentation ablation experiment results

Method	Single object		Multi objects	
	Total time / s	IoU / %	Total time / s	IoU / %
K-means	0.098	92.4	0.105	93.5
K-means+OTSU	0.258	96.3	0.324	96.8
OTSU	0.076	95.6	0.086	94.2
Improved OTSU	0.023	99.1	0.034	98.9

为了验证所提算法中引入的连通域和多采样约束的有效性,用改进的 OTSU 得到的图像作为输入,以测试多采样约束在 8 个不同场景中的效果。表 9 对比了有无连通域的实验结果,其中质心点偏移是预测坐标与真值的比率,无效域百分比是扫描连通域中错误区域的占比,CD 表示连通域,MSC 表示多采样约束。没有连通域模块会在目标提取过程中选择大量的无效区域(高达 84.3%),这将严重影响后续的位姿估

表 9 连通域消融实验结果

Table 9 Connected domain ablation experiment results

Method	Centroid shift / %	Invalid region percentage / %
Without CD	—	84.3
CD	3.21	3.42
CD+MSC	0.42	0.36

计。与直接标注连通域相比,在对图像扫描时引入多采样约束后,无效区域减少了 89.5%,质心点漂移现象减少,验证了多采样约束提高了区域选择的准确性,并优化了分区问题。

## 5 结 论

提出了基于 RGB-D 图像的 MinBRect 算法,用于单目标和多目标场景中的物体抓取。在目标检测阶段,通过求解质心点得到目标的位置,在姿态估计阶段,通过求解最小边界矩形得到旋转位姿角度。最后控制机械臂执行抓取任务。实验结果验证所提算法的速度和精度均有提升,可以完成多场景抓取任务。另外,所提算法减少了训练目标模型的工作量,在小样本及未知目标的检测上具有良好的泛化能力,为机械臂抓取提供研究思路。因此,所提算法为需要实时检测、计算资源有限、无预学习样本的工业结构化场景提供了算法支持。

## 参 考 文 献

- [1] 张磊, 徐孝彬, 曹晨飞, 等. 基于动态特征剔除的图像与点云融合的机器人位姿估计方法[J]. 中国激光, 2022, 49(6): 0610001.  
Zhang L, Xu X B, Cao C F, et al. Robot pose estimation method based on image and point cloud fusion with dynamic feature elimination[J]. Chinese Journal of

- Lasers, 2022, 49(6): 0610001.
- [2] Maturana D, Scherer S. VoxNet: a 3D Convolutional Neural Network for real-time object recognition[C]//2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), September 28-October 2, 2015, Hamburg, Germany. New York: IEEE Press, 2015: 922-928.
- [3] Charles R Q, Hao S, Mo K C, et al. PointNet: deep learning on point sets for 3D classification and segmentation [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 77-85.
- [4] Shi S S, Wang X G, Li H S. PointRCNN: 3D object proposal generation and detection from point cloud[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 770-779.
- [5] 徐冠宇, 董洪伟, 钱军浩, 等. 基于点对特征的散乱堆放物体的位姿估计算法[J]. 激光与光电子学进展, 2020, 57(18): 181508.
- Xu G Y, Dong H W, Qian J H, et al. Pose estimation algorithm for random bins based on point pair features[J]. Laser & Optoelectronics Progress, 2020, 57(18): 181508.
- [6] 张磊, 张天益, 姚兴田, 等. 机器人视觉中目标物的位姿估计改进[J]. 南通大学学报(自然科学版), 2021, 20(4): 38-45.
- Zhang L, Zhang T Y, Yao X T, et al. Improvement on pose estimation of an object in the robotic vision[J]. Journal of Nantong University (Natural Science Edition), 2021, 20(4): 38-45.
- [7] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [8] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.
- [9] 王凤随, 王启胜, 陈金刚, 等. 基于注意力机制和 Soft-NMS 的改进 Faster R-CNN 目标检测算法[J]. 激光与光电子学进展, 2021, 58(24): 2420001.
- Wang F S, Wang Q S, Chen J G, et al. Improved Faster R-CNN target detection algorithm based on attention mechanism and Soft-NMS[J]. Laser & Optoelectronics Progress, 2021, 58(24): 2420001.
- [10] 顾桂梅, 陈充, 余晓宁, 等. 基于改进 Faster R-CNN 的接触网管帽目标定位算法[J]. 激光与光电子学进展, 2022, 59(4): 0410009.
- Gu G M, Chen C, Yu X N, et al. Target location algorithm of contact network pipe cap based on improved Faster R-CNN[J]. Laser & Optoelectronics Progress, 2022, 59(4): 0410009.
- [11] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2021-05-06]. <https://arxiv.org/abs/1804.02767>.
- [12] Yu J M, Zhang W. Face mask wearing detection algorithm based on improved YOLO-v4[J]. Sensors, 2021, 21(9): 3263.
- [13] Yan B, Fan P, Lei X Y, et al. A real-time apple targets detection method for picking robot based on improved YOLOv5[J]. Remote Sensing, 2021, 13(9): 1619.
- [14] 崔海华, 姜涛, 杜坤鹏, 等. 基于深度学习位姿估计的多视结构光三维成像方法[J]. 光学学报, 2021, 41(17): 1712001.
- Cui H H, Jiang T, Du K P, et al. 3D imaging method for multi-view structured light measurement via deep learning pose estimation[J]. Acta Optica Sinica, 2021, 41(17): 1712001.
- [15] Hinterstoisser S, Lepetit V, Ilic S, et al. Model based training, detection and pose estimation of texture-less 3D objects in heavily cluttered scenes[M]//Lee K M, Matsushita Y, Rehg J M, et al. Computer vision-ACCV 2012. Lecture notes in computer science. Heidelberg: Springer, 2013, 7724: 548-562.
- [16] Sundermeyer M, Marton Z C, Durner M, et al. Implicit 3D orientation learning for 6d object detection from RGB images[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Heidelberg: Springer, 2018, 11210: 699-715.
- [17] He Y S, Sun W, Huang H B, et al. PVN3D: a deep point-wise 3D keypoints voting network for 6DoF pose estimation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11629-11638.
- [18] Dong Z K, Liu S C, Zhou T, et al. PPR-net: point-wise pose regression network for instance segmentation and 6D pose estimation in Bin-picking scenarios[C]//2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), November 3-8, 2019, Macao, China. New York: IEEE Press, 2019: 1773-1780.
- [19] Lin Y B, Zeng L, Dong Z K, et al. A vision-guided robotic grasping method for stacking scenes based on deep learning [C]//2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), October 11-13, 2019, Chongqing, China. New York: IEEE Press, 2019: 91-96.
- [20] 李喆雨, 丁坤, 张经纬, 等. 基于选择增强的光伏阵列遮挡分割研究[J]. 激光与光电子学进展, 2021, 58(24): 2410006.
- Li Z Y, Ding K, Zhang J W, et al. Selective enhancement-based shade segmentation of photovoltaic array[J]. Laser & Optoelectronics Progress, 2021, 58(24): 2410006.
- [21] 刘志成, 王殿伟, 刘颖, 等. 基于二维伽马函数的光照不均匀图像自适应校正算法[J]. 北京理工大学学报, 2016, 36(2): 191-196, 214.
- Liu Z C, Wang D W, Liu Y, et al. Adaptive adjustment algorithm for non-uniform illumination images based on 2D Gamma function[J]. Transactions of Beijing Institute of Technology, 2016, 36(2): 191-196, 214.
- [22] Chelishchev P, Sørby K. Estimation of minimum volume of bounding box for geometrical metrology[J]. International Journal of Metrology and Quality Engineering, 2020, 11: 9.