

# 基于 Kinect 的机器人抓取系统研究

黄玲涛 王彬 倪涛 缪海峰 李亚男

(吉林大学机械与航空航天工程学院, 长春 130022)

**摘要:** 智能抓取搬运机器人能够高效、可靠地完成各种搬运任务,降低工作人员的劳动强度,精准的物体定位是机器人执行搬运任务的基础。本文研究了基于 Kinect 的机器人抓取系统,可实现物体的类别检测、物体定位及机器人抓取任务。抓取系统由3个子系统(物体检测系统、物体定位系统及机器人抓取系统)组成。首先利用 Kinect 采集的物体图像信息训练单次多盒检测(Single multi-box detection, SSD)模型,然后根据 SSD 模型对物体的类别进行检测,得到物体在图像中的边框,并获取边框中物体像素坐标和深度,接着通过 Kinect 相机手眼标定法将像素坐标和深度转换到机器人基坐标系中,实现物体的定位,最后通过机器人逆运动学求解关节角,驱动机器人运动完成抓取搬运任务。对机器人进行了物体的定位和抓取实验,实验结果表明,物体的定位误差较小,物体抓取搬运实验的平均成功率达到97%,满足物体的抓取搬运需求。

**关键词:** 机器人; Kinect 相机; 物体检测; 物体抓取; 手眼标定

**中图分类号:** TP242      **文献标识码:** A      **文章编号:** 1000-1298(2019)01-0390-10

## Research on Robotic Grasping System Based on Kinect Camera

HUANG Lingtao WANG Bin NI Tao MIAO Haifeng LI Ya'nán

(College of Mechanical and Aerospace Engineering, Jilin University, Changchun 130022, China)

**Abstract:** The intelligent grasping robot can efficiently and reliably perform various handling tasks, reducing the labor intensity of the staff, and accurate object positioning was the basis for the robot to perform the handling task. A robotic grasping system was studied by using Kinect sensor. The robotic grasping system consisted of three subsystems (object detection system, object positioning system and robot motion system). The image information of the object acquired by Kinect sensor was firstly used to train a single multi-box detection (SSD) model, and then the object's category was detected according to the SSD model, the border of the object in the image and pixel coordinate and depth value of the border were obtained. In order to obtain the mapping between the coordinate information of the object in the three-dimensional space and the pixel information in the image, the camera can be calibrated by ZHANG Zhengyou calibration method to obtain the intrinsic parameters, extrinsic parameters and distortion parameters. The robot system and the vision system were connected by the Kinect camera hand-eye calibration, and then the three-dimensional coordinates of the object in the robot base coordinate system were obtained by coordinate transformation to realize the object positioning. Finally, the robot inverse kinematics was introduced to solve the joint angle, and the robot motion was driven to complete the grasping and conveying task. Object's positioning and conveying experiments were conducted. The mean absolute error of the object in  $x$  and  $y$  directions was 5.2 mm and 2.8 mm, respectively. The mean absolute error of object height was 4.5 mm. The success rate of the object grabbing and conveying experiment was 97%. The experimental results showed that the robotic grasping system proposed was valid to perform object detection, object positioning and conveying task.

**Key words:** robotics; Kinect camera; object detection; object grasping; hand-eye calibration

收稿日期: 2018-09-29 修回日期: 2018-11-02

基金项目: 国家自然科学基金项目(51575219)

作者简介: 黄玲涛(1979—),男,讲师,博士,主要从事机器人控制和主从控制及力反馈技术研究, E-mail: hlt@jlu.edu.cn

通信作者: 倪涛(1978—),男,教授,博士生导师,主要从事机器人技术和机器视觉处理研究, E-mail: nitao@jlu.edu.cn

## 0 引言

随着人工智能的快速发展,机器人技术的研究水平得到了飞速的提高,计算机视觉和机器人技术的结合成为当前的研究热点。受 HUBEL 等<sup>[1]</sup>的生物研究启发,卷积神经网络(Convolutional neural network, CNN)<sup>[2-4]</sup>被提出,并在物体识别、物体检测<sup>[5]</sup>等领域取得了成功。通过计算机视觉技术<sup>[5]</sup>对物体进行识别,然后控制机器人实现抓取成为研究的一个重要领域,该技术的实现可以摆脱传统的示教模式,应对不同的作业环境,并且能够大大降低人工成本,提高经济效益。

目前,在物体检测方面,牛津大学计算机视觉组(Visual geometry group, VGG)提出的 VGG 模型(Visual geometry group network, VGGNet)<sup>[6]</sup>成功地构建了 VGGNet-11、GGNet-13、VGGNet-16 和 VGGNet-19 等卷积神经网络模型。VGGNet-16 被用在深度学习 SSD 模型<sup>[7-9]</sup>中,并能够使模型获得较高的运行速度和检测精度。

在物体定位方面,为了得到物体在三维空间的坐标信息和在图像中的像素信息之间的映射,可以通过文献[10-11]方法对相机进行标定,获得其内参、外参、畸变参数等。文献[12-15]通过手眼标定将机器人系统和视觉系统连接起来,然后通过坐标变换得到物体在机器人基坐标系中的三维坐标,完成机器人的抓取任务<sup>[16-18]</sup>。

在机器人抓取方面,LENZ 等<sup>[19]</sup>将深度学习方法引入到机器人抓取中。LEVINE 等<sup>[20]</sup>收集机器人抓取实验数据用于训练神经网络模型,该模型通过卷积神经网络预测手爪成功抓取物体的概率,并完成抓取动作。

本文提出将 Kinect 摄像头作为视觉采集器,对物体识别及机器人的抓取搬运进行研究,利用 SSD 模型对物体类型进行实时识别,通过张正友标定法<sup>[10-11]</sup>获取 Kinect 内外参并计算物体位置信息,最后控制机器人各关节,以完成机器人对物体的抓取搬运任务。

## 1 系统组成

物体抓取系统结构如图 1 所示, Kinect 相机与机器人分别固定在工作环境中,相机固定在工作台,其不随机械手一起运动。物体抓取系统包括物体检测系统、物体定位系统和机器人抓取系统。物体检测系统是基于 SSD 模型的深度学习方法,通过对 20 个目标物体采样、标记和训练后形成适合本系统的物体检测模型。物体识别模型在 Kinect 相机获得

工作场景的实时彩色图像后,识别出目标物体并显示物体类别和边框,同时输出物体在像素坐标系中的像素坐标。物体定位系统获得物体在图像中的像素位置以及对应的深度值后,通过手眼标定方法,得到物体在机器人基坐标下的位置。机器人抓取系统在得到目标物体坐标值后,通过逆运动学求解得到机器人 6 个关节需要到达的位置,控制机器人及其机械手爪实现抓取搬运动作。

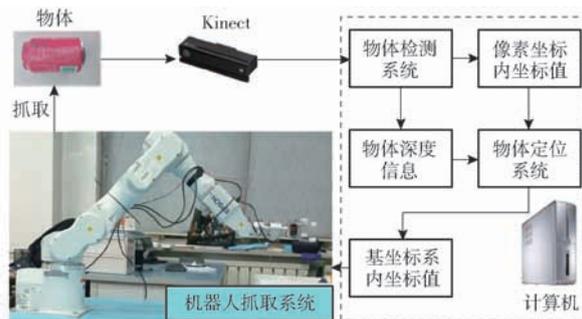


图 1 物体抓取系统结构示意图

Fig. 1 Schematic of object grasping system

## 2 物体检测系统

Kinect 相机在采集物体不同位置和姿态下的彩色图像后,对这些图像中的物体类别及其在图像中的像素坐标进行标注,然后将彩色图像及其标注信息作为训练样本,训练并生成 SSD 的物体检测模型,实现目标物体检测。

### 2.1 物体图像采集及增强

以 20 种物体(易拉罐、矿泉水瓶和纸盒等)作为实验的目标物体(图 2)。复杂环境背景的图像会对模型训练、评价和使用产生很大的干扰,为了减少这些干扰,机器人工作区域设置为浅蓝色。

用于训练模型的物体图像信息不仅需要物体处在不同位置和姿态状况下,而且还需要多个物体同时在一幅图像中。为了充分扩充数据集,对原图像作顺时针旋转  $10^\circ$ 、逆时针旋转  $10^\circ$  处理,并对原图像以及旋转图像进行高斯噪声处理和模糊处理,即一幅原图像能够得到 8 幅扩充图像(图 3)。实验共采集了 449 幅物体图像,数据集共  $449 \times 9 = 4\,041$  幅图像。

### 2.2 物体图像标注、转换及 SSD 模型

基于 SSD 模型的物体检测技术属于监督学习,所以需要获得图像中所有目标物体的标注信息,包括物体类别 id 和物体边框像素位置。

当所有的彩色图像均标注完毕后,需要将 70% 的图像和对应标注文件作为训练数据集,剩下 30% 作为评价数据集。虽然 Fast R-CNN 模型平均精度比 SSD 模型高一些,但是运行耗时却比 SSD 模型高很多,基本无法满足视频实时处理的要求,因此本文



图2 实验所用的20种物体

Fig.2 20 objects used in experiment

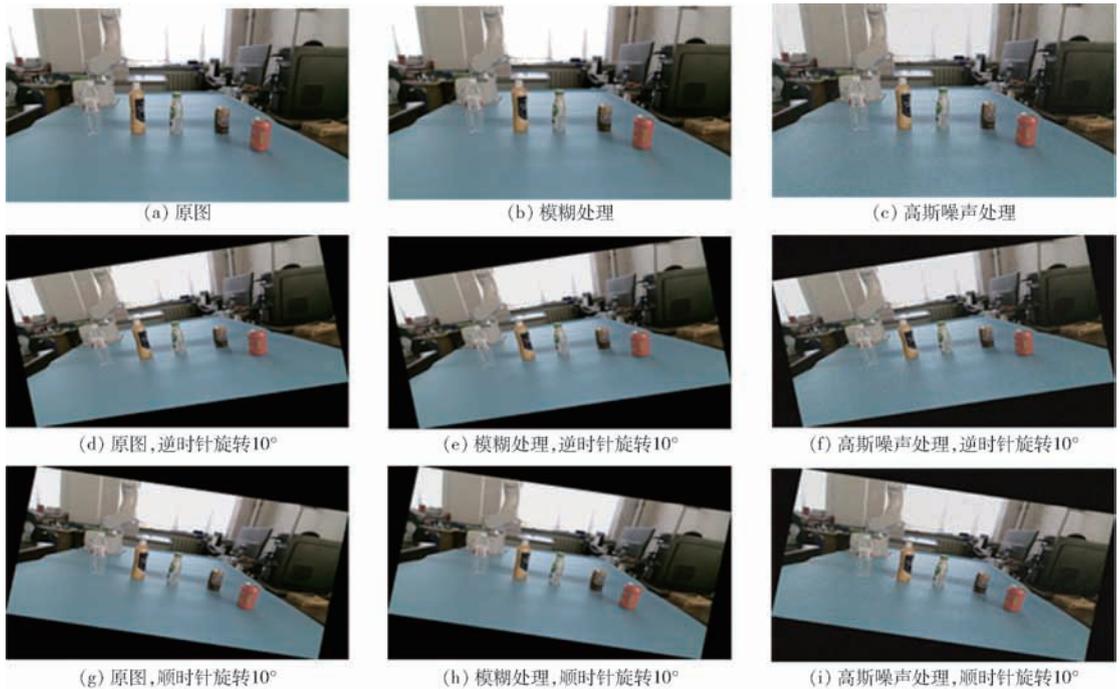


图3 原图的增强效果图

Fig.3 Enhancement effect pictures of original image

使用 SSD 训练模型, 训练模型的学习率设置为 0.0004, 批量大小设置为 1, 训练次数设置为 40000。随着训练次数的增加, 损失函数值随之减小。SSD 训练完成后, 使用评价数据集测试所得到的 SSD 模型, 其检测精度为 81.56%, 可以用来作为检测物体的模型。

### 3 机器人抓取系统

#### 3.1 机器人

使用 EPSON A901S 型六自由度机器人实现对物体的抓取, 其各坐标系如图 4 所示。机器人与实

验台连接处为机器人基坐标系 ( $\{0\}$ ), 关节 1~3 分别作为第  $\{1\} \sim \{3\}$  坐标系。由于此机器人的结构特点, 以及为了使机器人拥有逆运动学封闭解, 第  $\{4\} \sim \{6\}$  坐标系置于同一位置, 即关节 4 处。此机器人的特点是前 3 个关节控制机器人末端的位置, 后 3 个关节控制机器人末端的姿态。此后用  $\{i\}$  表示第  $i$  坐标系。

#### 3.2 机械手爪

机械手爪 (图 5a) 通过连接法兰固定在机器人末端, 用于抓取目标物体。机械手爪由数字舵机驱动完成手爪的张合动作, 并由 Arduino UNO 单片机

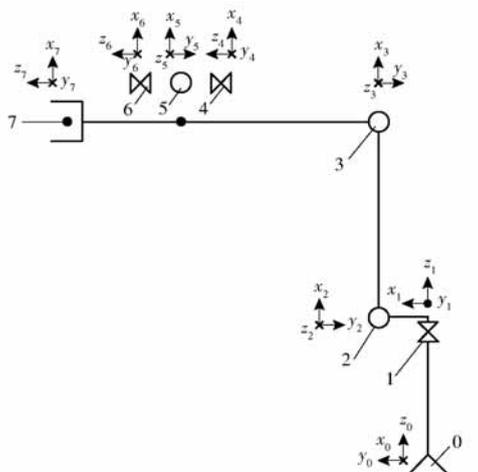


图 4 机器人坐标系简图

Fig. 4 Sketch of robot coordinate system

0. 基坐标系 1~7. 关节坐标系

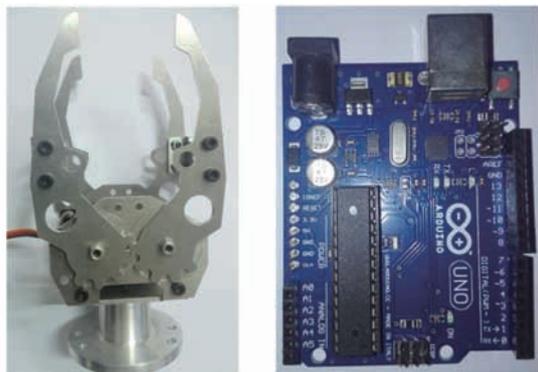


图 5 机械手爪和 Arduino UNO 单片机

Fig. 5 Mechanical gripper and Arduino UNO microcontroller

(图 5b) 编程控制。

### 3.3 机器人 D-H 参数

根据 EPSON 六自由度机器人、连接器和机械手爪的尺寸以及各个关节坐标系的位置和姿态, 机器人 D-H 参数如表 1 所示。其中  $\varphi_1 \sim \varphi_6$  为机器人关节 1~6 的关节角, 该关节角为 EPSON 生产厂家定义, 与 D-H 参数中的  $\theta_i$  关系见表 1。

表 1 EPSON 六自由度机器人 D-H 参数

Tab. 1 D-H parameters of EPSON six-degree-of-freedom robot

$i$	$\alpha_{i-1}/(^{\circ})$	$a_{i-1}/m$	$d_i/m$	$\theta_i/(^{\circ})$	$\varphi_i/(^{\circ})$
1	0	0	0.320	$90 + \varphi_1$	$[-170, 170]$
2	90	0.100	0	$90 + \varphi_2$	$[-160, 65]$
3	0	0.400	0	$\varphi_3$	$[-51, 225]$
4	90	0	0.400	$-\varphi_4$	$[-200, 200]$
5	-90	0	0	$\varphi_5$	$[-135, 135]$
6	90	0	0	$-\varphi_6$	$[-360, 360]$
7	0	0	0.276	0	0

### 3.4 机器人正运动学

假设  $\{i\}$  到  $\{i-1\}$  的旋转矩阵为  ${}^{i-1}\mathbf{R}$ ,  $\{i\}$  原点

在  $\{i-1\}$  的平移向量为  ${}^{i-1}\mathbf{P}$ , 则  $\{i\}$  到  $\{i-1\}$  的齐次变换矩阵  ${}^{i-1}\mathbf{T} = \begin{bmatrix} {}^{i-1}\mathbf{R} & {}^{i-1}\mathbf{P} \\ 0 & 1 \end{bmatrix}$ 。根据机器人 D-H 参数(表 1), 可得齐次变换矩阵  ${}^{i-1}\mathbf{T} (i=1, 2, \dots, 7)$ 。

### 3.5 机器人逆运动学

如图 4 所示, 机械手爪末端中心位置为第  $\{7\}$  坐标系的原点位置。当机械手爪安装在机器人末端时, 为实现手爪到达指定的位置和姿态, 必须对机器人的逆运动学求解。由于 EPSON A901S 机器人的结构特点, 即第 4~6 坐标系交于一点满足 PIEPER 准则, 此机器人具有封闭解。

由于  $\{7\}$  相对于机器人基坐标系的位置和姿态已知, 为  $(p_{x7}, p_{y7}, p_{z7}, \theta_{x7}, \theta_{y7}, \theta_{z7})^T$ , 故  $\{7\}$  相对于基坐标系的齐次变换矩阵为

$${}^0_7\mathbf{T} = \begin{bmatrix} c_z c_y & c_z s_y s_x - s_z c_x & c_z s_y c_x + s_z s_x & p_{x7} \\ s_z c_y & s_z s_y s_x + c_z c_x & s_z s_y c_x - c_z s_x & p_{y7} \\ -s_y & c_y s_x & c_y c_x & p_{z7} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

其中  $s$  表示  $\sin$ ,  $c$  表示  $\cos$ , 其下标  $x$  表示  $\theta_{x7}$ ,  $y$  表示  $\theta_{y7}$ ,  $z$  表示  $\theta_{z7}$ 。

$\{7\}$  相对于  $\{6\}$  仅发生了平移, 姿态没有改变, 所以  $\{6\}$  相对于  $\{0\}$  的姿态与  $\{7\}$  相对于  $\{0\}$  的姿态相同, 故  $\{6\}$  的齐次变换矩阵为

$${}^0_6\mathbf{T} = {}^0_7\mathbf{T}_7\mathbf{T}^{-1} = \begin{bmatrix} * & * & * & p_{x7} - 0.276(c_z s_y c_x + s_z s_x) \\ * & * & * & p_{y7} - 0.276(s_z s_y c_x - c_z s_x) \\ * & * & * & p_{z7} - 0.276c_y c_x \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

其中,  $*$  表示未计算的值。由齐次变换矩阵的特点可知, 矩阵第 4 列前 3 个值分别是  $\{6\}$  在基坐标系中的位置。令  $\{6\}$  在基坐标系中的位姿为  $(p_x, p_y, p_z, \theta_x, \theta_y, \theta_z)^T$ , 其中有

$$\begin{cases} p_x = p_{x7} - 0.276(c_z s_y c_x + s_z s_x) \\ p_y = p_{y7} - 0.276(s_z s_y c_x - c_z s_x) \\ p_z = p_{z7} - 0.276c_y c_x \\ \theta_x = \theta_{x7} \\ \theta_y = \theta_{y7} \\ \theta_z = \theta_{z7} \end{cases}$$

然后计算  ${}^0_1\mathbf{T}^{-1} {}^0_6\mathbf{T} = {}^1_2\mathbf{T}_2^3\mathbf{T}_3^4\mathbf{T}_4^5\mathbf{T}_5^6\mathbf{T}$  左右两边的矩阵, 并且令左右两边  $4 \times 4$  矩阵各个元素对应相等, 则可以

$$p_x \cos \varphi_1 + p_y \sin \varphi_1 = 0 \quad (1)$$

$$\frac{p_z - 0.32}{0.4} = \cos \varphi_2 + \sin(\varphi_2 + \varphi_3) \quad (2)$$

$$\frac{p_x \cos\varphi_1 - p_x \sin\varphi_1 - 0.1}{0.4} = \cos(\varphi_2 + \varphi_3) - \sin\varphi_2 \tag{3}$$

### 3.5.1 关节角 $\varphi_1$

根据式(1)可得

$$\varphi_1 = \arctan\left(-\frac{p_x}{p_y}\right) \tag{4}$$

$\varphi_1$  存在多解, 其为  $\varphi_{11}$  和  $\varphi_{12}$ , 当  $\varphi_1 > 0$  时,  $\varphi_{11} = \varphi_1$ ,  $\varphi_{12} = \varphi_{11} - \pi$ 。当  $\varphi_1 \leq 0$  时,  $\varphi_{12} = \varphi_1$ ,  $\varphi_{11} = \varphi_{12} + \pi$ 。

### 3.5.2 关节角 $\varphi_3$

对式(2)和式(3)等号两边同时平方再相加, 可解得

$$\varphi_3 = \arcsin\left(\frac{m^2 + n^2}{2} - 1\right) \tag{5}$$

其中

$$m = \frac{p_z - 0.32}{0.4}$$

$$n = \frac{p_x \cos\varphi_1 - p_x \sin\varphi_1 - 0.1}{0.4}$$

每一个  $\varphi_1$  对应的  $\varphi_3$  有两个值, 即  $\varphi_{31} = \varphi_3$  和  $\varphi_{32} = -\varphi_3$ , 因为  $\varphi_1$  本身有两个解, 所以  $\varphi_3$  有 4 个值。

### 3.5.3 关节角 $\varphi_2$

将  $\varphi_1$  和  $\varphi_3$  代入式(2)和式(3)中, 求解得

$$\varphi_2 = \arccos\left(\frac{n \cos\varphi_3}{2(\sin\varphi_3 + 1)} + \frac{m}{2}\right) \tag{6}$$

其中,  $\varphi_2$  的正负号与  $\sin\varphi_2$  的正负号相同。每一个  $\varphi_3$  唯一对应一个  $\varphi_2$  的值, 所以  $\varphi_2$  也有 4 个值。

### 3.5.4 筛选最优关节角 $\varphi_1$ 、 $\varphi_2$ 和 $\varphi_3$

由于  $\varphi_1$  有两个解, 即  $\varphi_{11}$  和  $\varphi_{12}$ , 每个  $\varphi_1$  对应两个  $\varphi_3$ , 每个  $\varphi_3$  唯一对应一个  $\varphi_2$ , 所以共有 4 组解:  $\{\varphi_{11}, \varphi_{21}, \varphi_{31}\}$ 、 $\{\varphi_{11}, \varphi_{22}, \varphi_{32}\}$ 、 $\{\varphi_{12}, \varphi_{23}, \varphi_{33}\}$ 、 $\{\varphi_{12}, \varphi_{24}, \varphi_{34}\}$ 。虽然前 3 个关节角有 4 组解, 但是由于机器人各个关节均有旋转角范围, 故每个关节角必须在此范围之内, 关节角范围如表 1 所示。若此时只剩下一组解, 则此组关节角为最终的前 3 个关节角; 如仍然剩下多组解, 则表示这些角均可使机器人达到期望位姿, 此时可依据转动幅度最小原则筛选出最优关节角。得到的最优关节角  $\varphi_1$ 、 $\varphi_2$  和  $\varphi_3$  将用于计算后 3 个关节角  $\varphi_4$ 、 $\varphi_5$  和  $\varphi_6$ 。

### 3.5.5 关节角 $\varphi_5$

因为  $\varphi_1$ 、 $\varphi_2$ 、 $\varphi_3$  已知, 故可求  ${}^0_3T^{-1} {}^0_6T = {}^3_6T$ , 其中  ${}^3_6T$  第 2 行第 3 列元素仅与  $\varphi_5$  有关, 两边矩阵对应项相等可解

$$\varphi_5 = \arccos(c_{2+3}(s_{z-1}c_x s_y - c_{z-1}s_x) + s_{2+3}c_x c_y) \tag{7}$$

其中下标“2+3”表示  $\varphi_2 + \varphi_3$ , “z-1”表示  $\theta_{x7} - \varphi_1$ 。 $\varphi_5$  有两个解, 即  $\varphi_{51} = \varphi_5$  和  $\varphi_{52} = -\varphi_5$ 。

### 3.5.6 关节角 $\varphi_4$

${}^3_6T$  第 1 行第 3 列元素仅与  $\varphi_4$  有关, 当  $\varphi_5 \neq 0$  时, 可解得

$$\varphi_4 = \begin{cases} \left| \arccos \frac{g}{\sin\varphi_5} \right| & (\sin\varphi_4 > 0) \\ - \left| \arccos \frac{g}{\sin\varphi_5} \right| & (\sin\varphi_4 \leq 0) \end{cases} \tag{8}$$

其中  $g = s_{2+3}c_x(c_{z-1} - s_{z-1}c_y) + c_{2+3}c_x c_y$ 。每一个  $\varphi_5$  唯一对应一个  $\varphi_4$ , 因  $\varphi_5$  有 2 个, 故  $\varphi_4$  有 2 个值。

### 3.5.7 关节角 $\varphi_6$

${}^3_6T$  的第 2 行第 1 列元素仅与  $\varphi_6$  有关,  $\varphi_5 \neq 0$  时, 可解得

$$\varphi_6 = \begin{cases} \left| \arccos \frac{h}{\sin\varphi_5} \right| & (\sin\varphi_6 > 0) \\ - \left| \arccos \frac{h}{\sin\varphi_5} \right| & (\sin\varphi_6 \leq 0) \end{cases} \tag{9}$$

其中  $h = s_{2+3}(c_1^2 c_y + s_1^2 s_y) - c_{2+3}s_{z-1}c_y$ 。每一个  $\varphi_5$  对应一个  $\varphi_6$ , 因为存在 2 个  $\varphi_5$ , 故  $\varphi_6$  有 2 个值。由式(8)、(9)可知, 当  $\varphi_5 = 0$  时, 关节 5 处于奇异点位置, 相应的雅可比矩阵不可逆,  $\varphi_4$  和  $\varphi_6$  无解, 此时不能使用公式求解关节角  $\varphi_4$  和  $\varphi_6$ 。此时, 本文令  $\varphi_5$  为一个极小值 ( $0.0001^\circ$ ), 将其代入式(8)、(9)中求解  $\varphi_4$  和  $\varphi_6$ 。

### 3.5.8 筛选最优关节角 $\varphi_4$ 、 $\varphi_5$ 和 $\varphi_6$

由于  $\varphi_5$  有 2 个解, 即  $\varphi_{51}$  和  $\varphi_{52}$ , 每个  $\varphi_5$  唯一对应一个  $\varphi_4$  和  $\varphi_6$ , 所以共有 2 组解:  $\{\varphi_{41}, \varphi_{51}, \varphi_{61}\}$ 、 $\{\varphi_{42}, \varphi_{52}, \varphi_{62}\}$ 。首先需要判断 2 组解是否均在关节旋转角范围内, 若排除 1 组解则剩下 1 组解为最终的后 3 个关节角; 如仍然剩下 2 组解, 则表示这些角均可使机器人达到期望位姿, 此时与筛选前 3 个关节角的方法一样筛选后 3 个关节角。

## 4 基于 Kinect 相机的物体定位

物体检测系统识别物体时, 能生产物体图像边框, 边框 4 点连线交点为中心点, 并作为机器人搬运时的抓取点。文中的物体定位是针对物体中心点的定位。由中心点的像素坐标系坐标及其深度值, 求其在相机坐标系的位置, 然后通过坐标变换将其相机坐标位置转换到  $\{0\}$  系中的位置 ( $X_0, Y_0, Z_0$ ), 通过逆运动学求解, 控制机器人各关节角, 以实现机器人对目标物体的抓取。

### 4.1 彩色图像映射到深度空间

Kinect 深度相机与 RGB 彩色相机安装位置不同, 所以观察视角也不同, 其彩色图像的分辨率是 1920 像素  $\times$  1080 像素, 深度图像分辨率是 512 像

素  $\times 424$  像素, 彩色图像上的每一个像素 ( $1\ 920 \times 1\ 080 = 2\ 073\ 600$ ) 并不能在深度图像上找到一一对应的深度 ( $512 \times 424 = 217\ 088$ )。Kinect for Windows SDK 中提供一个将彩色帧映射到深度空间的函数 (`MapColorFrameToDepthSpaceUsingIntPtr`)。该函数会得到一个长度与彩色帧像素个数相同的一维数组 ( $P[2073600]$ ), 数组中的元素表示原深度图的坐标位置, 此坐标位置的值表示相应的深度值, 将这些深度依次存放在与数组  $P$  相同长度的一维数组 ( $D[2073600]$ ) 中。最后将数组  $D$  转换成  $1\ 920 \times 1\ 080$  的矩阵  $M$ 。彩色图像的每一个像素都能在矩阵  $M$  的相应位置得到其深度  $Z_c$ ,  $Z_c$  被用来求物体中心点在  $\{0\}$  中的位置信息 ( $X_0, Y_0, Z_0$ )。

#### 4.2 Kinect 相机标定

利用张正友标定法对 Kinect 相机进行标定。影响标定误差的主要因素为标定板精度及标定照片数量。此外还受光照强度、标定图片标中定板面积占比、板面平整度等方面影响<sup>[21]</sup>。实验标定板采用的是  $11 \times 8$  黑白相间的平整正方形棋盘格 (图 6), 每个正方形边长为  $30\text{ mm}$ , 其精度为  $\pm 0.005\text{ mm}$ 。

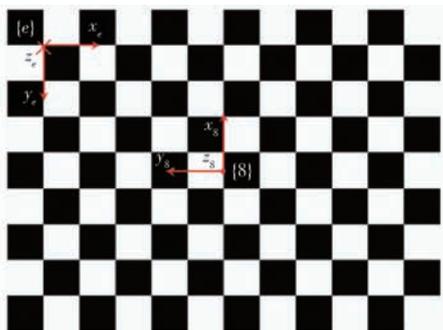


图 6 标定板及其物理坐标

Fig. 6 Calibration board and its physical coordinates

机器人末端去掉抓手, 安装标定板后, 标定板中心为  $\{8\}$  坐标系, 标定板坐标系为  $\{e\}$ , 位于标定板的左上方。标定板中心点与六维力传感器中心点对齐后, 将其固定。在光照充足的情况, 移动机器人末端位置和姿态使标定板处于不同位姿, 通过 Kinect 获得 20 幅标定图像, 并确保每幅图像中标定板面积占比超过 50%。

##### 4.2.1 相机坐标系转换到图像坐标系

从相机坐标系  $\{c\}$  转换到图像坐标系是将三维点转换到二维点, 假设目标物体在  $\{c\}$  的坐标为  $(X_c, Y_c, Z_c)$ , 在图像坐标系上的投影坐标为  $(x, y)$ 。如图 7 所示, 根据相似三角形法, 整理后  $\{c\}$  到图像坐标系的转换关系为

$$Z_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (10)$$

式中  $f$ ——焦距

$[x \ y \ 1]^T$ ——归一化后的图像物理坐标

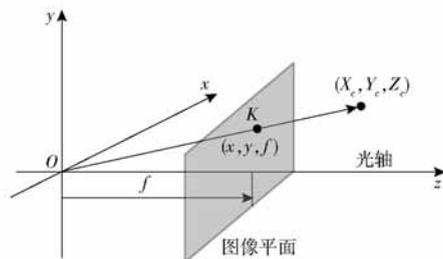


图 7 物体在相机坐标系和图像坐标系投影示意图

Fig. 7 Projection of an object in camera coordinate

##### 4.2.2 相机坐标系转换到像素坐标系

点  $K$  在图像坐标系中的坐标为  $(x, y)$ , 在像素坐标系中的坐标为  $(u, v)$ 。图像坐标系转换到像素坐标系的转换关系为

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{Z_c} \begin{bmatrix} \frac{x}{dx} + u_0 \\ \frac{y}{dy} + v_0 \\ 1 \end{bmatrix} = \frac{1}{Z_c} \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \frac{1}{Z_c} \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (11)$$

式中  $dx, dy$ ——图像中单个像素在  $x, y$  方向上分别占据的长度单位数

$u_0, v_0$ ——图像中心像素坐标与图像原点像素坐标间相差的横向、纵向像素数

式(11)可以改写成

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \mathbf{KX}_c \quad (12)$$

式中  $\mathbf{K}$ ——相机内参, 为  $3 \times 4$  矩阵

$\mathbf{X}_c$ ——目标物体在  $\{c\}$  的坐标

##### 4.2.3 像素坐标系转换到基坐标系

从世界坐标系转换到  $\{c\}$  是将三维点转换到另一个三维点。在相机标定过程中, 将世界坐标系与  $\{e\}$  设置为重合。令目标物体在世界坐标系 ( $\{e\}$ ) 的坐标值为  $\mathbf{X}_e = (X_e, Y_e, Z_e)^T$ ,  $\{e\}$  到  $\{c\}$  的旋转矩阵为  $\mathbf{R}$ , 平移向量为  $\mathbf{t}$ , 则  $\{e\}$  到  $\{c\}$  的关系为

$$\mathbf{X}_c = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{O}^T & 1 \end{bmatrix} \begin{bmatrix} X_e \\ Y_e \\ Z_e \\ 1 \end{bmatrix} \quad (13)$$

将式(13)代入式(12)整理可得

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = KX_c = K \begin{bmatrix} R & t \\ O^T & 1 \end{bmatrix} \begin{bmatrix} X_e \\ Y_e \\ Z_e \\ 1 \end{bmatrix} = K_e^c T \begin{bmatrix} X_e \\ Y_e \\ Z_e \\ 1 \end{bmatrix} \quad (14)$$

其中  ${}^c T = \begin{bmatrix} R & t \\ O^T & 1 \end{bmatrix}$

式中  ${}^c T$ —— $\{e\}$  到  $\{c\}$  的变换矩阵  
 $O$ —— $3 \times 1$  零向量

由于标定板固定在机器人末端,标定板与机器人之间的位姿关系是确定的,故  $\{e\}$  中的一点可以表示为

$$\begin{bmatrix} X_e \\ Y_e \\ Z_e \\ 1 \end{bmatrix} = {}^0 T \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \\ 1 \end{bmatrix} \quad (15)$$

结合式(14),可得像素坐标系到  $\{0\}$  的转换为

$$\begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \\ 1 \end{bmatrix} = {}^0 T^{-1} \begin{bmatrix} X_e \\ Y_e \\ Z_e \\ 1 \end{bmatrix} = {}^0 T (K_e^c T)^{-1} Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = Z_{c_e}^0 T_c^e T K^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = Z_{c_e}^0 T K^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad (16)$$

式中  ${}^0 T$ —— $\{c\}$  到  $\{0\}$  的齐次变换矩阵,也被称为相机外参,为  $4 \times 4$  矩阵

当 Kinect 相机安装完成之后,其位置和姿态固定不变,机器人基坐标系也是不变的,所以  ${}^0 T$  可以唯一确定。其表达式为

$${}^0 T = {}^0 T_e^c T = {}^0 T_1^1 T_2^2 T_3^3 T_4^4 T_5^5 T_6^6 T_7^7 T_e^e T \begin{bmatrix} R & t \\ O^T & 1 \end{bmatrix}^{-1}$$

假设  $\{c\}$  在  $\{0\}$  中的位姿表示为  ${}^0 P_c = (p_{cx}, p_{cy}, p_{cz}, \theta_{cx}, \theta_{cy}, \theta_{cz})^T$ , 可得  $\{c\}$  到  $\{0\}$  的齐次变换矩阵为

$${}^0 T' = \begin{bmatrix} c_z c_y & c_z s_y s_x - s_z c_x & c_z s_y c_x + s_z s_x & p_{cx} \\ s_z c_y & s_z s_y s_x + c_z c_x & s_z s_y c_x - c_z s_x & p_{cy} \\ -s_y & c_y s_x & c_y c_x & p_{cz} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

由于  $\{c\}$  到  $\{0\}$  的齐次变换矩阵唯一,故  ${}^0 T = {}^0 T'$ 。左右两边矩阵对应元素相对,可求出  ${}^0 P_c$ 。再由  ${}^0 P_c$  计算  ${}^0 T$ ,并将其代入式(16)中,可得像素坐标系到机器人基坐标系的转换关系。

#### 4.2.4 标定结果

Kinect 在获得 20 幅棋盘格 RGB 图像后,将其转换成灰度图;通过亚像素角点来获得棋盘格角点像素点坐标。因为世界坐标系建立在棋盘格平面

上,所以棋盘格角点在世界坐标系的  $Z$  轴坐标为零,棋盘格内所有角点在世界坐标系中的坐标值是已知的。由 20 幅棋盘格图像获得的 Kinect 相机内参矩阵为

$$K = \begin{bmatrix} 1059.4 & 0 & 965.24 \\ 0 & 1068.3 & 523.77 \\ 0 & 0 & 1 \end{bmatrix}$$

20 幅标定图像计算得相机位姿平均值  ${}^0 \bar{P} = (1.2639, 0.354, 0.3067, -1.68484, 0.0072, 1.5343)^T$ , 由  ${}^0 \bar{P}$  计算得到

$${}^0 \bar{T} = \begin{bmatrix} 0.0164 & -0.0183 & -0.9997 & 1.2773 \\ 0.9998 & 0.0128 & 0.0162 & 0.5737 \\ 0.0125 & -0.9998 & 0.0185 & 0.3205 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

## 5 实验

### 5.1 物体检测实验

将 20 种目标物体依次随机摆放在工作台不同位置上,使用 SSD 模型检测这些物体,SSD 模型会将每一个物体的名称及其准确率显示在物体边框左上方,并用不同的颜色表示不同的物体类别(图 8)。

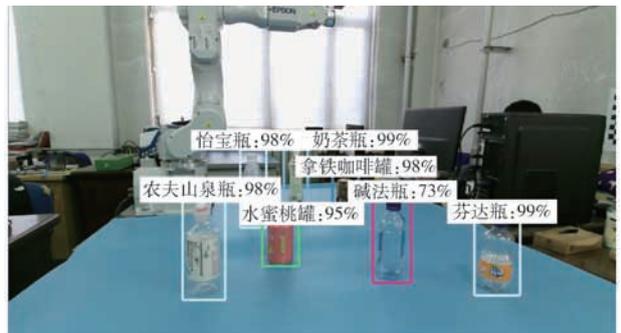


图 8 物体检测结果

Fig. 8 Object detection results

根据边框位置可以确定物体中心点在图像中的准确位置,如图 9 所示,图 9a 为深度图像(经彩色图像映射深度空间处理后图像),中心点代表物体的深度值,图 9b 为彩色图像,中心点坐标表示物体像素坐标。图中物体“百事可乐罐”的深度  $Z_c$  为 1218.75 mm,物体的像素坐标得到后,根据式(16)可计算出物体在基坐标系的准确位置。

### 5.2 物体定位实验

实验平台上选择 4 个位置,其坐标分别是  $(0.8, 0.2)$ 、 $(0.75, 0.35)$ 、 $(0.5, 0.7)$ 、 $(0.6, 0.6)$ ,然后将 20 个物体分别放置在此 4 个位置上(黑色“+”字标记所在位置),计算得物体位置如图 10 中粉色、蓝色、红色和绿色所示的“x”。20 组数据所得的平均值及方差如表 2 所示,物体定位  $x$  向平均

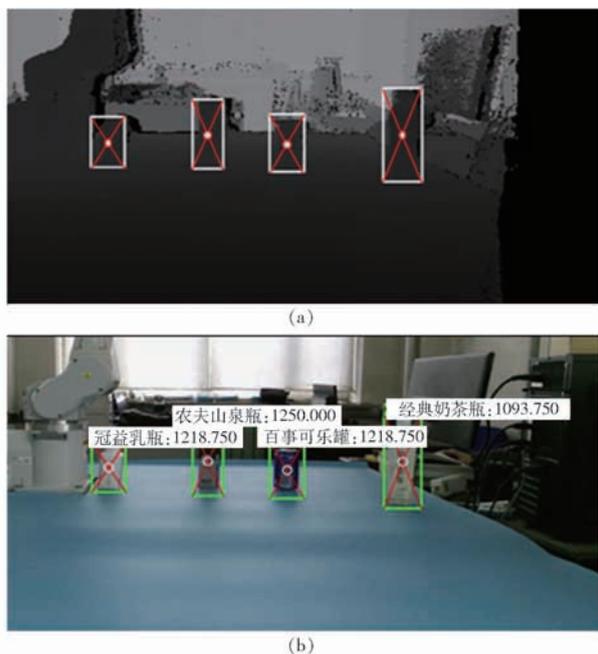


图 9 目标物体的像素坐标和深度

Fig. 9 Pixel coordinates and depth values of target objects

绝对误差为 5.2 mm,  $y$  向为 2.8 mm。物体高度 ( $H_r$ ) 及测量值的平均高度 ( $H_m$ ) 如表 3 所示, 物体高度的平均绝对误差为 4.5 mm。由测量结果可知, 物体定位能够比较准确地测算出物体在平台的位置, 并且绝对误差较小, 能够满足抓取任务要求。

表 3 物体高度及其测量平均值

Tab. 3 Object height and measured average

序号	$H_r$	$H_m$									
1	115.5	116.9	6	227.7	221.1	11	137.5	139.6	16	209.3	200.4
2	127.5	133.3	7	227.5	225.4	12	180.0	167.8	17	182.2	193.3
3	133.2	138.1	8	178.6	174.5	13	162.2	159.7	18	235.5	230.7
4	115.5	114.6	9	220.5	213.8	14	196.2	191.1	19	115.6	117.2
5	218.5	211.1	10	146.5	147.3	15	232.1	233.6	20	121.4	122.6

### 5.3 物体抓取实验

为减少背景环境的干扰, 实验台的背景设置成浅蓝色, 并在光线充足的条件下进行物体抓取实验。在实验平台上选择 6 个区域 (图 10), 分别将 20 个物体依次摆放在 6 个区域内, 机器人在每个区域内抓取物体 5 次, 每个物体被抓取次数为  $6 \times 5 = 30$  次; 实验总次数为  $20 \times 6 \times 5 = 600$  次。实验抓取动作流程如图 11 所示。

抓取过程中出现失败抓取情况, 机器人不重新尝试再次抓取该物体, 操控者需将该目标物体移出抓取区, 继续后续其他物体抓取实验。即使出现抓取失败物体碰倒其他物体的情况, 只需将被碰倒的物体重新放回原位, SSD 模型会对物体进行再次识别, 程序会重新计算该物体中心点的位置。

表 4 表示 6 个区域中 20 种物体被抓取实验的

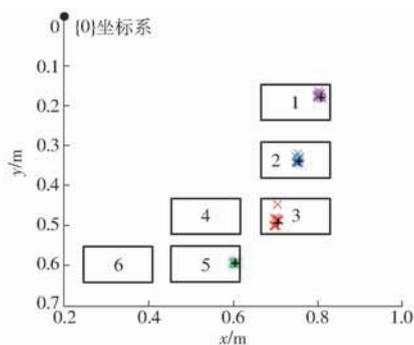


图 10 工作台物体定位实验结果

Fig. 10 Experimental results of object positioning on workbench

表 2 物体坐标测量平均值及其方差

Tab. 2 Mean and variance of object's positioning

mm			
	坐标	平均值	方差
$x$ 向	800	794.5	3.5
	750	745.4	3.1
	500	496.3	11.7
	600	600.4	2.6
$y$ 向	200	496.3	11.7
	350	350.3	6.4
	700	695.7	3.8
	600	600.1	3.0

成功率, 其中区域 2 和区域 4 的成功率最高, 达到 99%, 区域 6 的成功率最低, 但是也能够达到 95%。表 5 表示每个目标物体被抓取的成功率, 其中 10 个目标物体被抓取成功率为 100%, 物体 13 的成功率最低, 为 83%。物体抓取失败的主要原因是该物体的中间部分形状不规则, 为凹陷弯曲, Kinect 在测量此类物体时其深度信息存在较大的误差, 最终导致定位系统所得到的物体位置也存在较大误差。本实验的 600 次抓取中成功次数为 582 次, 总成功率为 97%, 证明了通过深度学习检测物体并将位置映射到机器人基坐标系的方法在机器人抓取任务中是有效的。

### 6 结束语

设计了一个基于 Kinect 相机的机器人抓取系

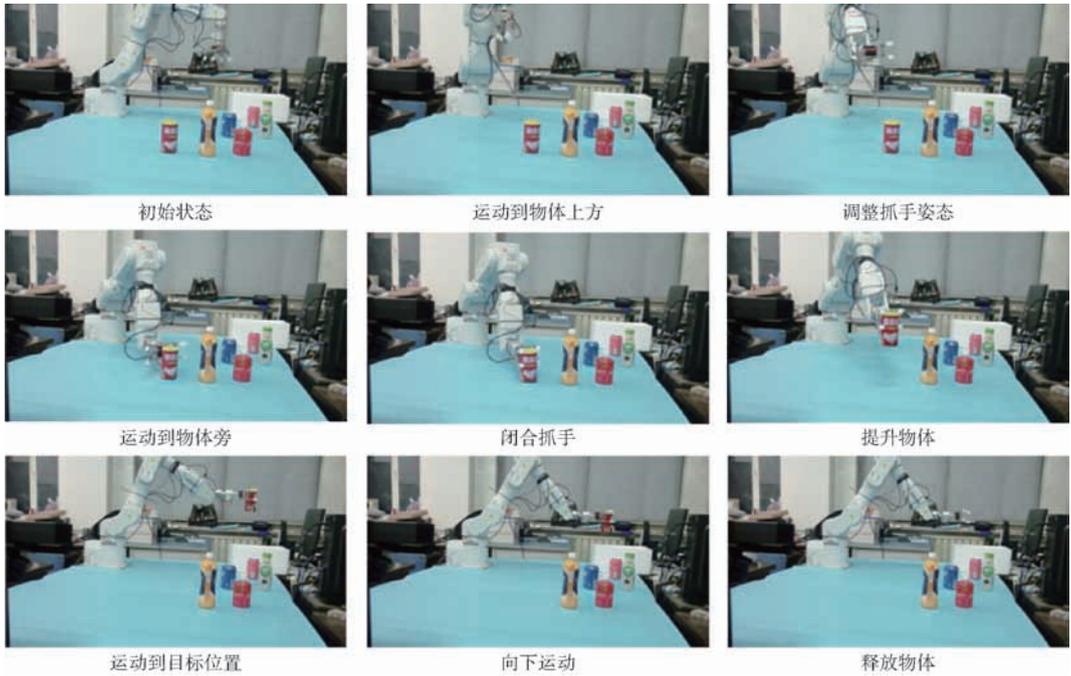


图 11 机器人抓取物体动作流程

Fig. 11 Action diagram of object grasping experiment

表 4 物体区域抓取实验的成功率

Tab. 4 Success rate of object grasping experiment

项目	inix regions					
	区域					
	1	2	3	4	5	6
成功率/%	96	99	96	99	97	95

表 5 物体抓取实验的成功率

Tab. 5 Success rate of object grasping experiment

目标物体编号	1	2	3	4	5	6	7	8	9	10
成功率/%	96	100	100	100	96	93	97	100	100	96
目标物体编号	11	12	13	14	15	16	17	18	19	20
成功率/%	93	100	83	93	100	100	100	93	97	100

统,利用该系统可实时对作业环境内的物体进行识别检查,准确计算物体的定位信息。对所构建的系统分别进行了物体检查、定位及抓取实验。物体在  $x$  向平均绝对误差为 5.2 mm,  $y$  向平均绝对误差为 2.8 mm,物体高度的平均绝对误差为 4.5 mm,误差

较小,满足物体抓取时的定位要求,物体抓取实验的成功率达到 97%。实验结果表明,所构建的机器人系统针对不同目标物体能实时、可靠地完成物体抓取搬运任务。

参 考 文 献

[1] HUBEL D H, WIESEL T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex[J]. The Journal of Physiology, 1962, 160(1): 106 - 154.

[2] LENG B, LIU Y, YU K, et al. 3D object understanding with 3D convolutional neural networks[J]. Information Sciences, 2016, 366: 188 - 201.

[3] KRIZHEYSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]// Advances in Neural Information Processing Systems, 2012: 1097 - 1105.

[4] SZEGEDY C, TOSHEV A, ERHAN D. Deep neural networks for object detection[C]// Advances in Neural Information Processing Systems, 2013: 2553 - 2561.

[5] SERMANET P, EIGEN D, ZHANG X, et al. OverFeat: integrated recognition, localization and detection using convolutional networks[EB/OL]. [2013 - 12 - 21] <https://arxiv.org/abs/1312.6229>.

[6] SIMONVAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. [2014 - 09 - 04] <https://arxiv.org/abs/1409.1556>.

[7] LIU W, ANGUELOY D, ERHAN D, et al. Ssd: Single shot multibox detector[C]// European Conference on Computer vision, 2016: 21 - 37.

[8] HOLSCHNEIDER M, KRONLAND-MARTINET R, MORLET J, et al. A real-time algorithm for signal analysis with the help of the wavelet transform[C]// JEAN-MICHEL Combes. Wavelets. Time-Frequency Methods and Phase Space. Berlin: Springer, 1990: 286 - 297.

- [9] HUANG J, RATHOD V, SUN C, et al. Speed/accuracy trade-offs for modern convolutional object detectors [C] // IEEE CVPR, 2017.
- [10] ZHANG Z. A flexible new technique for camera calibration [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11): 1330 - 1334.
- [11] ZHANG Z. Flexible camera calibration by viewing a plane from unknown orientations [C] // Proceedings of the 7th IEEE International Conference on Computer Vision, 1999: 666 - 673.
- [12] 张洪瑶, 李论, 周波, 等. 机器人手眼标定及其精度分析的研究 [J]. 组合机床与自动化加工技术, 2018(1): 69 - 72.  
ZHANG Hongyao, LI Lun, ZHOU Bo, et al. Research on hand-eye calibration of robots and its accuracy analysis [J]. Modular Machine Tool & Automatic Manufacturing Technique, 2018(1): 69 - 72. (in Chinese)
- [13] HORAUD R, DORNAIKA F. Hand-eye calibration [J]. The International Journal of Robotics Research, 1995, 14(3): 195 - 210.
- [14] ZHUANG H, ROTH Z S, SUDHAKAR R. Simultaneous robot/world and tool/flange calibration by solving homogeneous transformation equations of the form  $AX = YB$  [J]. IEEE Transactions on Robotics and Automation, 1994, 10(4): 549 - 554.
- [15] DORNAIKA F, HORAUD R. Simultaneous robot-world and hand-eye calibration [J]. IEEE Transactions on Robotics and Automation, 1998, 14(4): 617 - 622.
- [16] 唐新星, 倪涛, 何丽鹏, 等. 基于立体视觉的遥操作工程机器人自主作业系统 [J/OL]. 农业机械学报, 2012, 43(10): 224 - 228.  
[http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?flag=1&file\\_no=20121040&journal\\_id=jcsam](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20121040&journal_id=jcsam). DOI: 10.6041/j.issn.1000-1298.2012.10.040.  
TANG Xinxing, NI Tao, HE Lipeng, et al. Autonomous task control system of construction tele-robot based on stereo vision [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2012, 43(10): 224 - 228. (in Chinese)
- [17] 李宪华, 盛蕊, 张雷刚, 等. 六自由度模块化机器人手臂奇异构型分析 [J/OL]. 农业机械学报, 2017, 48(7): 376 - 382.  
[http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?flag=1&file\\_no=20170748&journal\\_id=jcsam](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20170748&journal_id=jcsam). DOI: 10.6041/j.issn.1000-1298.2017.07.048.  
LI Xianhua, SHENG Rui, ZHANG Leigang, et al. Singular configuration analysis of 6-DOF modular manipulator [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(7): 376 - 382. (in Chinese)
- [18] MILLER A T, AIKEN P K. Graspit! a versatile simulator for robotic grasping [J]. IEEE Robotics & Automation Magazine, 2004, 11(4): 110 - 122
- [19] LENZ I, LEE H, SAXENA A. Deep learning for detecting robotic grasps [J]. The International Journal of Robotics Research, 2015, 34(4-5): 705 - 724.
- [20] LEVINE S, PASTOR P, KRIZHEVSKY A, et al. Learning hand-eye coordination for robotic grasping with large-scale data collection [C] // International Symposium on Experimental Robotics, 2016: 173 - 184.
- [21] 支健辉, 董新民, 孔星炜, 等. 相机标定的外界影响因素分析 [J]. 应用光学, 2014, 35(2): 286 - 291.  
ZHI Jianhui, DONG Xinmin, KONG Xingwei, et al. Analysis of external influence factors in camera calibration [J]. Journal of Applied Optics, 2014, 35(2): 286 - 291. (in Chinese)