

基于抓取点生成网络 与视觉伺服的未知物体抓取算法研究

王泽飞

院（系）： 航天学院 专 业： 自动化
学 号： 1181140118 指导教师： 高会军

2022 年 6 月 x 日

哈爾濱工業大學

畢業設計（論文）

題 目 基於抓取點生成網絡與視
覺伺服的未知物體抓取算法研究

專 業 自動化

學 號 1181140118

學 生 王澤飛

指 導 教 師 高會軍

答 辯 日 期 2022 年 6 月 x 日

摘 要

摘要的字数（以汉字计），硕士学位论文一般为 500 ~ 1000 字，博士学位论文为 1000 ~ 2000 字，均以能将规定内容阐述清楚为原则，文字要精练，段落衔接要流畅。摘要页不需写出论文题目。英文摘要与中文摘要的内容应完全一致，在语法、用词上应准确无误，语言简练通顺。留学生的英文版博士学位论文中应有不少于 3000 字的“详细中文摘要”。

关键词是为了文献标引工作、用以表示全文主要内容信息的单词或术语。关键词不超过 5 个，每个关键词中间用分号分隔。（模板作者注：关键词分隔符不用考虑，模板会自动处理。英文关键词同理。）

关键词：TeX；L^ATeX；CJK；嗨！；thesis

Abstract

An abstract of a dissertation is a summary and extraction of research work and contributions. Included in an abstract should be description of research topic and research objective, brief introduction to methodology and research process, and summarization of conclusion and contributions of the research. An abstract should be characterized by independence and clarity and carry identical information with the dissertation. It should be such that the general idea and major contributions of the dissertation are conveyed without reading the dissertation.

An abstract should be concise and to the point. It is a misunderstanding to make an abstract an outline of the dissertation and words “the first chapter”, “the second chapter” and the like should be avoided in the abstract.

Key words are terms used in a dissertation for indexing, reflecting core information of the dissertation. An abstract may contain a maximum of 5 key words, with semi-colons used in between to separate one another.

Keywords: T_EX, L^AT_EX, CJK, template, thesis

物理量名称及符号表

表 1 国际单位制中具有专门名称的导出单位

量的名称	单位名称	单位符号	其它表示实例
频率	赫[兹]	Hz	s ⁻¹

目 录

摘 要	I
Abstract	II
物理量名称及符号表	III
第 1 章 绪论	1
1.1 课题背景及研究的目的和意义	1
1.2 机器人视觉伺服系统概述	1
1.3 基于视觉伺服的物体抓取发展现状	1
1.4 主要研究内容及章节安排	2
第 2 章 基于图像的视觉伺服（IBVS）理论研究	3
2.1 引言	3
2.2 IBVS 系统模型建立	3
2.2.1 系统坐标系建立	3
2.2.2 视觉模型建立	4
2.3 IBVS 仿真系统实现	5
2.3.1 IBVS 算法原理	5
2.3.2 机器人仿真模型搭建	6
2.3.3 曲线绘制与相机轨迹记录	7
2.4 IBVS 实际系统实现	8
2.4.1 IBVS 实物系统平台搭建	8
2.4.2 IBVS 实际运行	9
2.5 本章小结	10
第 3 章 伺服目标生成算法研究	11
3.1 引言	11
3.2 基于模型的点云识别与配准	11
3.2.1 方法陈述	11
3.2.2 算法实现	12
3.2.3 方法小结	12
3.3 抓取点生成网络（GG-CNN）	13

3.3.1 方法陈述	13
3.3.2 算法实现	14
3.3.3 方法小结	16
3.4 本章小结.....	17
第 4 章 未知物体抓取算法实现	18
4.1 引言	18
4.2 机器人视觉伺服系统概述	18
第 5 章 IBVS 控制律优化	19
第 6 章 实验设计与验证	20
6.1 引言	20
6.2 实验设计.....	20
6.3 人机交互界面设计	21
6.4 实验验证.....	21
6.4.1 抓取成功率测试	21
6.4.2 抓取性能验证	22
6.5 本章小结.....	22
结 论	23
参考文献	24
哈尔滨工业大学本科毕业设计（论文）原创性声明	25
致 谢	26

第 1 章 绪论

1.1 课题背景及研究的目的和意义

要不看

1.2 机器人视觉伺服系统概述

分 IBVS、PBVS 和混合控制 hvs。IBVS 的基础原理从最原始的公式、解耦、图像矩到现在借助神经网络获直接端到端控制。绕不开当前特征与目标特征的提取的过程，但是方法越来越具有泛化性。

1.3 基于视觉伺服的物体抓取发展现状

抓取方法上：经典的方法当然是人为的制作特征点或者标志，辅助视觉伺服。但是泛化性是不行的，需要避免人为的制作特征，让程序自主提取特征并生成目标。直接把整张图当作特征是后来的发展方向之一，这样必须要求目标与当前相似，伺服范围太小。神经网络兴起后，方法变得百花齐放。如现在不断发展的抓取合成（grasp synthesis），分经验法和深度法；神经网络自主生成目标图像但还是那个问题，这样必须要求目标与当前相似，伺服范围太小；手到眼系统，神经网络估计机械臂末端和目标的相对位姿，伺服后期会出现遮挡现象，一般还是要配合眼在手系统一起使用，涉及到多数据融合问题，非常复杂。通过渲染引擎获得目标图像，神经网络自动提取特征并匹配特征。大部分的方法无法回避需要线下制作目标图像的问题，对于一个完全未知的目标，如何在线上就能实时生成目标并执行抓取是对抓取未知目标任务的研究重点。伺服性能上：就算拥有一个合适的抓取方法，也需要有合适的控制律。低鲁棒性的伺服控制律，无法在应对各种位姿、形态的抓取对象中保证同样的抓取性能，这会大大降低最终的抓取成功率和伺服响应速度。为满足泛化性的需求，近几年有很多人投入到视觉伺服控制律的研究中，然后介绍各种控制方法。如何设计一个适合于当前抓取方法的控制律，是抓取未知目标任务的又一难点。

归功于近年深度网络的快速发展，抓取合成（grasp synthesis）的方法在这几年发展迅速。

1.4 主要研究内容及章节安排

介绍本文都要干些啥，IBVS 相对于 PBVS 的优势，eye-to-hand 系统本文共分为六章，章节内容如下第一章：第二章：。。。

第2章 基于图像的视觉伺服（IBVS）理论研究

2.1 引言

绪论中介绍了 IBVS 相对于 PBVS 的优势：1. 伺服精度不依赖于相机外参，深度鲁棒性强；2. 直接得到指令，不需要轨迹规划，更易于实现实时闭环控制。因此认为 IBVS 更适合于未知位姿、形体的目标的抓取任务。IBVS 作为本算法研究的基石之一，尤其需要十分严谨合理的模型建立、公式原理分析、完善的仿真系统和实物环境搭建，这会为之后的工作减少不少麻烦。本章中除了完成上述基础性工作，还在实物上成功运行了简单的基于特征点交互矩阵的 IBVS。

2.2 IBVS 系统模型建立

2.2.1 系统坐标系建立

对于 IBVS 系统，最需要关注的点有三个：机械臂末端、相机和目标。为了后续仿真程序实现和问题分析需要，建立系统的坐标系用于表述它们的位置。如图2-1所示：

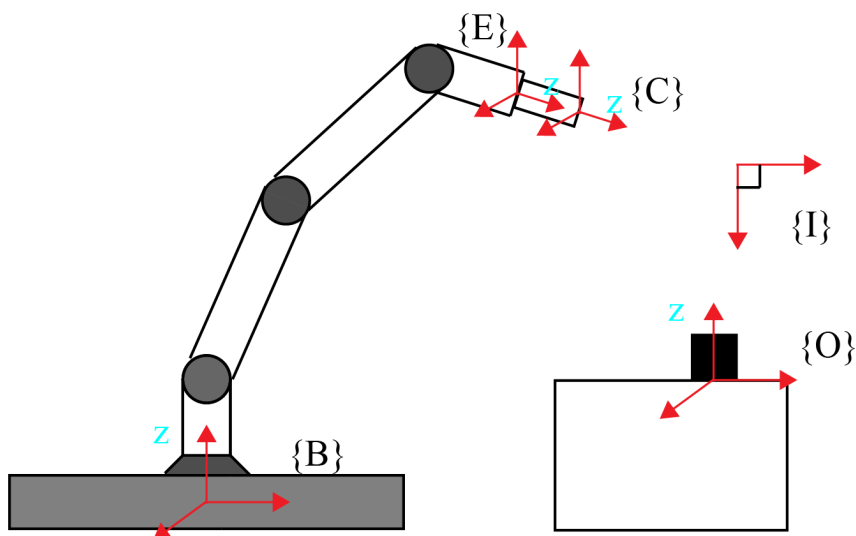


图 2-1 系统坐标系建立示意图

该示意图中， $\{O\}$ 、 $\{E\}$ 和 $\{C\}$ 分别表示物体坐标系、机器人末端坐标系，eye-to-hand 系统中的相机坐标系。为了便于阐述坐标变换公式，用 $\{B\}$ 、 $\{I\}$ 和 $\{CI\}$ 表示机器人基坐标系、图像坐标系和像素坐标系，本研究中所说的基坐标

系和世界坐标系是一个意思。将使用以下符号表示各个相对位姿变换： ${}^B T_O$ 表示目标 $\{O\}$ 相对于基坐标系 $\{B\}$ 的坐标变换； ${}^B T_E$ 表示机器人末端 $\{E\}$ 相对于基坐标系 $\{B\}$ 的变换。对机械臂末端使用的速度指令是在这个变化下进行的； ${}^C T_O$ 表示目标 $\{O\}$ 相对于相机坐标系 $\{C\}$ 的坐标变换； ${}^E T_C$ 表示相机 $\{C\}$ 相对于末端坐标系 $\{E\}$ 的坐标变换。一般情况下 IBVS 的伺服结果是相机正对目标，而真正抓取还是要依赖末端位置，所以这个变换是必要的； ${}^C T_I$ 表示图像 $\{I\}$ 相对于相机坐标系 $\{C\}$ 的坐标变换。特征初始是在图像中获取的，需要这个变换使特征位置描述变成 IBVS 需要的形式^[1]。

2.2.2 视觉模型建立

IBVS 不断地由特征偏差驱动着运行，而对特征的描述需要在 ${}^C T_O$ 下进行。在图像中获取的特征需要经过图2-2所示的坐标系变换才能真正为 IBVS 所用：



图 2-2 特征变换过程图

不例外地使用针孔模型描述从像素坐标系到机器人基坐标系中物体的映射, 这张图引自这篇文献^[1], 这张针孔模型示意图十分典型。

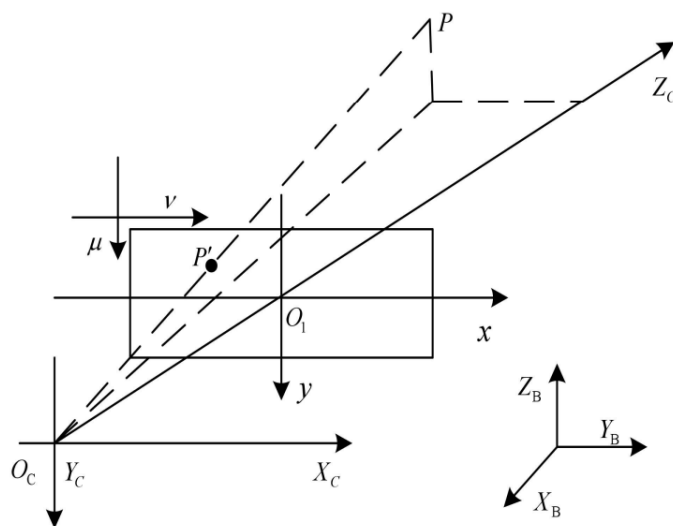


图 2-3 视觉坐标系建立示意图

图2-3中用 $X_C O_C Y_C$ 描述 $\{C\}$, $x O_I y$ 描述 $\{I\}$, P 表示 $\{C\}$ 中的目标点, P' 表示成像过程中投影到 $\{I\}$ 中的目标点。因为深度的存在, 成像在二维像素坐标系中的图像所对应的目标可以是无穷多种情况, 为了统一坐标变换形式, 令目标深度 Z_C 为单位 1, 在相机坐标系 $\{C\}$ 和像素坐标系 $\{CI\}$ 中加上了一个过渡

的图像坐标系 $\{I\}$ 。

相机内参由出厂地所给出，它包括相机的焦距 f ，相机放缩因子 f_x 和 f_y ，它们的单位为毫米；偏移量 c_x 和 c_y ，单位为像素，但是是浮点类型。由于 Realsense D435i 内置去畸变 API，就不考虑畸变因素了。使用 $[u, v]^T$ 表示像素坐标系下的目标点位置， $[X_C, Y_C, Z_C]^T$ 表示相机坐标系下的目标点位置， Z_C 为相机深度，可以得到它们之间的关系：

$$Z_C \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix} \quad (2-1)$$

在已知相机内参的情况下，IBVS 中所需要的特征位置描述就转换为找到特征对应的 $\{CI\}$ 中的位置 $[u, v]^T$ 和深度相机测得对应点的深度 Z_C 。

2.3 IBVS 仿真系统实现

2.3.1 IBVS 算法原理

在本实验中所说的 IBVS 是基于特征点交互矩阵实现的^[2]。它的基本思想是将特征点偏差通过交互矩阵（也叫图像雅可比矩阵）映射为末端速度指令。为了表述简洁，符号所代表的意思如下： $s^* = [x^*, y^*]^T$ 表示 $\{C\}$ 中对应的期望特征点坐标。 $s = [x, y]^T$ 表示 $\{C\}$ 中对应的当前特征点坐标。 $v_c = [v_x, v_y, v_z, \omega_x, \omega_y, \omega_z]^T$ 表示 $\{B\}$ 中相机的速度，其中相机包含质心线速度，和绕三个轴的角速度。由于相机和机械臂末端是固连且位置上接近的，所以它们的速度认为是一致的。使用交互矩阵 L_c 建立当前特征随时间变化率与相机位姿随时间变化率的关系：

$$\dot{s} = L_c v_c \quad (2-2)$$

一般情况下，期望特征是不随时间改变的，或者变化甚微（在本研究中就是如此，所以进行近似），式2-2可以如方程组2-3中第一排表达式改写。另外，认为特征偏差随时间呈指数变化是合理的，因为它收敛快速且平滑^[2]，于是可以得到方程组2-3：

$$\begin{cases} (\dot{s} - \dot{s}^*) = \dot{e} = L_c v_c \\ L_e = L_c \\ \dot{e} = -\lambda e \end{cases} \quad (2-3)$$

其中 \dot{e} 为当前特征偏差随时间变化率， λ 为比例系数， e 为当前特征偏差。

通过对交互矩阵求广义逆，由方程组2-3可得到：

$$\begin{cases} v_c = -\lambda L_e^+ e \\ L_e^+ = (L_e^T L_e)^{-1} L_e^T \end{cases} \quad (2-4)$$

其中 λL_e^+ 为交互矩阵广义逆。通过方程组2-4可以借助当前特征偏差求取机器人末端速度了。交互矩阵由特征点在图像中的位置及深度信息得到，每个点对应的交互矩阵如式2-5所示，若有多个点，公式中的交互矩阵就是每个点对应的交互矩阵在行方向的叠加。

$$L_e = \begin{bmatrix} \frac{-1}{Z_C} & 0 & \frac{X_C}{Z_C} & X_C Y_C - (1 + X_C^2) & Y_C \\ 0 & \frac{-1}{Z_C} & \frac{Y_C}{Z_C} & 1 + Y_C^2 & -X_C Y_C & -X_C \end{bmatrix} \quad (2-5)$$

2.3.2 机器人仿真模型搭建

基于为整个系统搭建的坐标系和不同系的坐标转换关系，借助 ROS 的 moveit 工具（由于 ROS2 的 moveit2 尚未开发成熟，使用 moveit 代替），为敬科公司提供的 JK 机器人搭建仿真模型。moveit 是一个开发的十分完善的工具包，不仅实现了机械结构的仿真，物理模型、碰撞体积和逆运动学都在包中相应地实现。本研究中，为了能更快地验证提出的算法，减少繁杂的处理，将把物体放到一个平整且颜色单一（在后续的研究中可以发现这些要求都不是必须的）的表面上。另外，相机一直保持俯视朝下，在 X、Y 轴方向的角度保持为 0° ，因此速度指令中 ω_x 和 ω_y 不论结果计算如何都给 0。仿真效果图如图2-4所示。

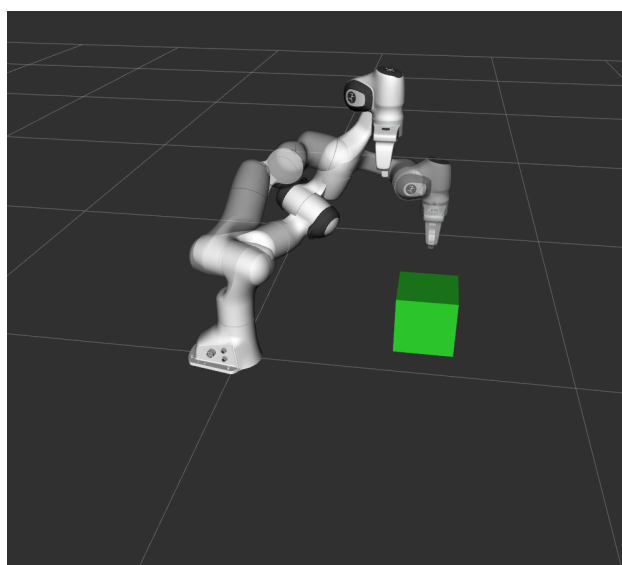


图 2-4 基于 moveit 机器人仿真模型实现

图2-5展现了整个系统最基础的控制方框图，在实验进行过程中会不断被改进，以应对实践中发生的各个问题。

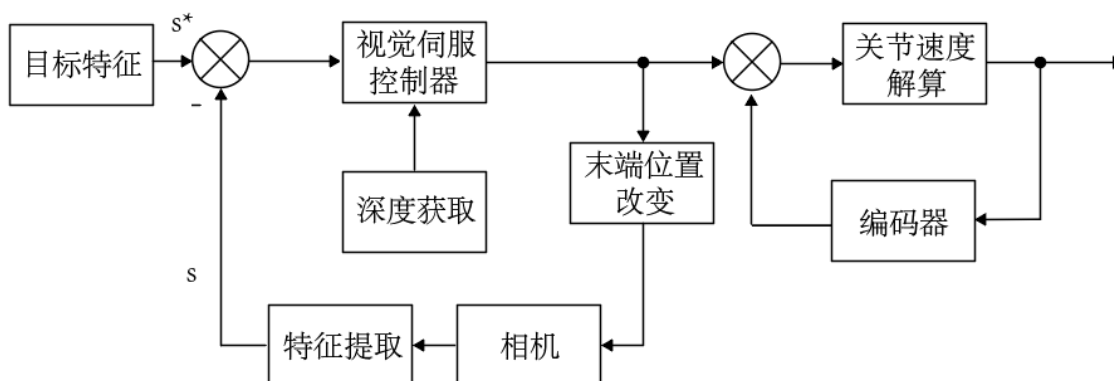


图 2-5 IBVS 基础控制流程方框图

深度获取的方式是会根据深度条件切换的，如图2-6所示。之所以加上这个切换，是因为 Realsense D435i 是基于结构光测量深度的，不可以测量过近距离的目标点。当相机在伺服末期十分靠近目标时，无法获得确切的目标点深度。所以在相机与目标点距离低于 20cm 时，会通过编码器读取末端下降的距离得到对应特征的深度，这在相机一直俯视向下时是可行的。



图 2-6 深度传感器切换示意图

开启仿真节点后，可以在另一个节点中与该节点建立连接并发送速度指令，仿真节点会因此响应，并进行移动。moveit 有自己的限位系统，在机器人进入奇异点或者超出移动范围时给予及时的警告，所以该机器人仿真模型多被用于对机器人是否进入奇异点的判断这样的定性分析，后文中真正的调参还是在实物上进行的。

2.3.3 曲线绘制与相机轨迹记录

曲线是分析问题非常重要的一环，所以仿真中应当有相应的曲线绘制。IBVS 本质是将特征偏差作为控制器输入而映射成速度指令的控制系统，所以研究中最关心的是点在于特征偏差和速度指令，它们将被分别绘制到两张图中。曲线图中时间单位为秒。关于特征偏差图：系统中定义特征偏差是相机坐标系

中被检测的特征点在 X、Y 方向的偏差，单位为米，该单位不被展现在曲线中，因为它的单位并不重要。关于末端速度指令图：为了与 JK 机器人需要的末端速度指令单位保持一致，所以线速度选取厘米每秒为单位，而角速度单位则为度每秒。

直观地展现相机的位移情况也是重要的，因为 IBVS 往往对机器人末端的运动轨迹十分不友好。如果当前的速度指令使机器人颤振，那么机器人已经进入了一个十分糟糕的姿态，通过分析相机的运动轨迹适当调节控制律参数也是非常好的解决方法。调用 VISP 库，对设定的参数进行视觉伺服仿真，实现的曲线绘制和相机轨迹绘制效果如图2-7所示。

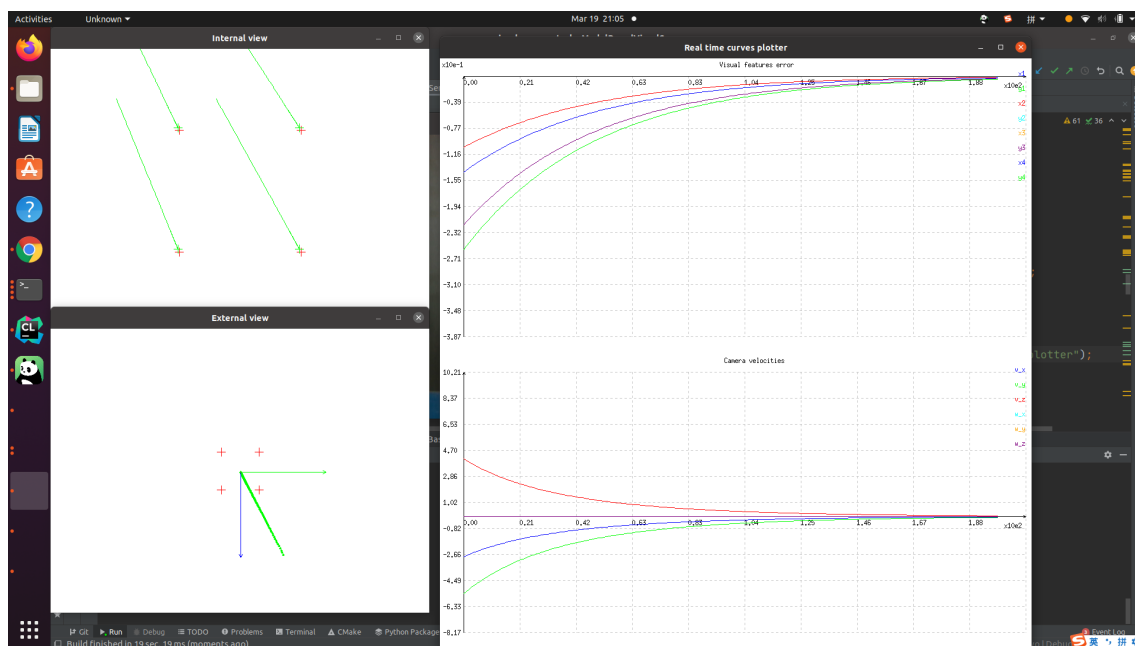


图 2-7 基于 VISP 视觉伺服仿真

提前在程序中设置好特征的位置，和与之对应的目标特征。将仿真得到的速度指令通过 ROS2 节点发布订阅机制传输给机器人仿真节点，机器人会相应地运动并使当前特征都到达目标特征处，从而到成到达目标位置处的目的。

2.4 IBVS 实际系统实现

2.4.1 IBVS 实物系统平台搭建

仿真终归只能用于定性分析。外界干扰、噪声多种多样，仿真中不可能把所有因素考虑进去。事实上，仿真跑出的结果往往十分顺滑，而实物中会反映很多处理不够细节的问题。我认为，IBVS 算法在实物上成功运行，研究才算真正的开始。实物运行环境包括 JK 机器人和装载它并固定它底座的台子；用于

承载目标物体的平台和目标; 平台上铺盖的一层漫反射效果好且为单一白色的纸; 机器人末端装配 Realsense D435i 深度相机 (夹具暂时未装配, 在正式夹取的时候会安装在末端)。之所以要铺一层纸, 除了保证平面平整且颜色单一以外, 还保证了深度相机不要因为丢失反射光导致获取无效数据。最终实物环境图如图2-8所示:

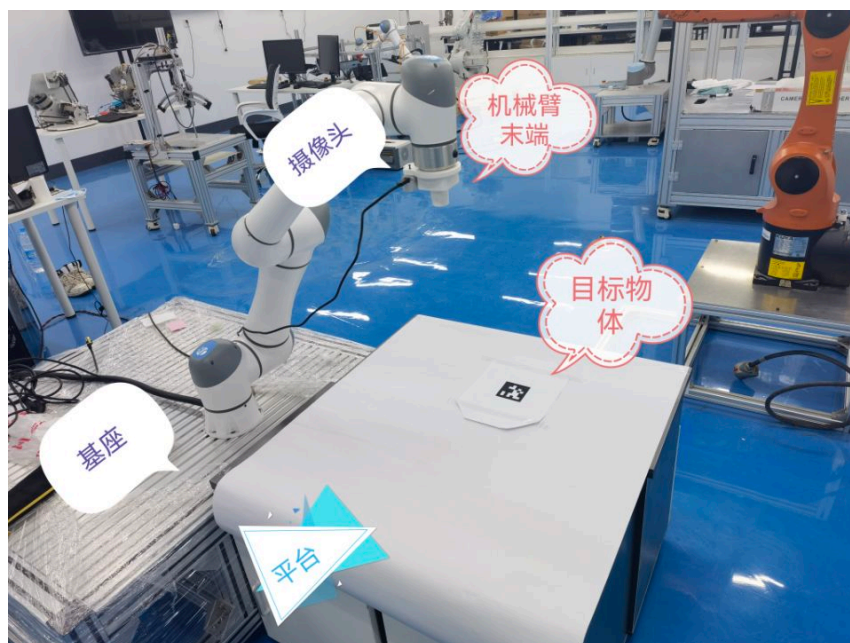


图 2-8 实物环境搭建展示图

2.4.2 IBVS 实际运行

为了能更快地验证 IBVS 算法, 在平台上贴了一张黑色方框码, 经过视觉二值化、边缘获取、多边形拟合等处理。实验进行前会将机械臂末端调到目标位置, 此时黑色方框会处于摄像头的中央, 记录此时的四个点为目标特征。将机器人末端初始位置调至远离黑色方框的位置, 距离目标位置的三维各个方向以及 Z 轴角度都有一定的偏差 ($\{B\}$ 中, $\Delta X = 0.3m$, $\Delta Y = 0.3m$, $\Delta Z = 0.5m$)。伺服过程中会不断捕获它的四个点作为特征, 并计算特征偏差, 最后映射成末端速度指令。伺服的成功证实了所实现的 IBVS 算法的正确性, 同时也正式踏入对未知物体视觉伺服抓取的研究领域中。2-9展示了伺服过程中机械臂末端速度指令和各特征点的偏差对应的曲线。

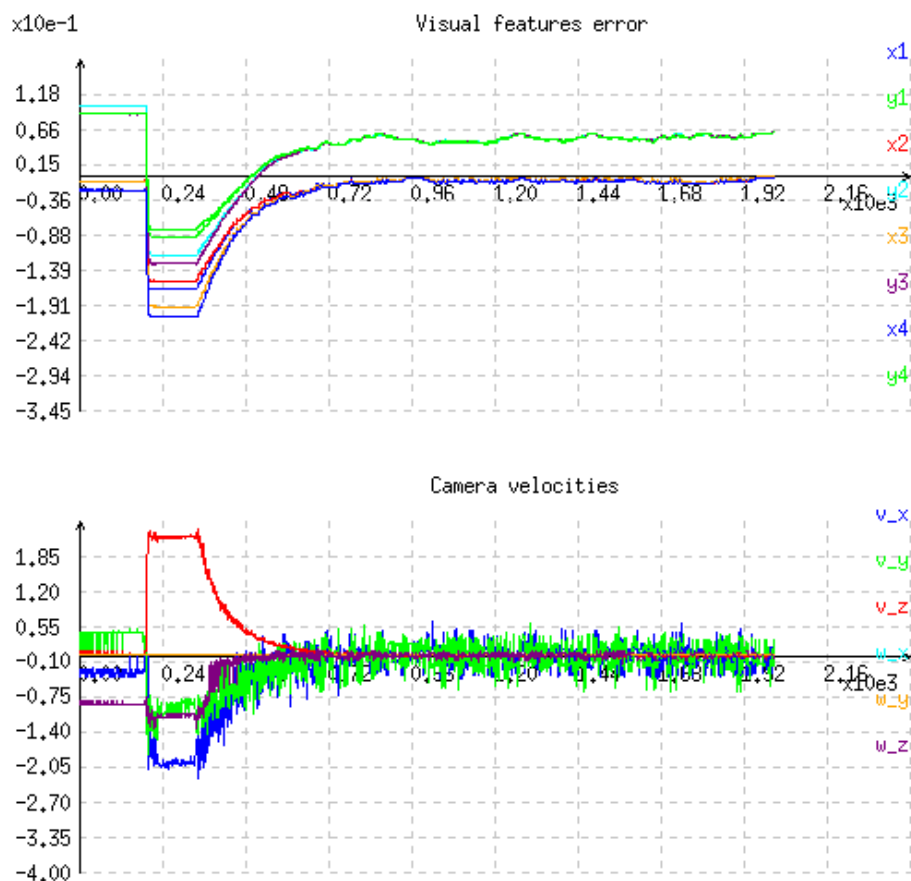


图 2-9 视觉伺服曲线绘制（初始）

2.5 本章小结

本章讲述了基于特征点交互矩阵的 IBVS 的原理。在算法实现前建立系统坐标系和视觉坐标系，这会使之后的坐标描述便捷许多。搭建了 IBVS 的仿真运行环境，便于后续问题分析。搭建了实物运行环境，并成功运行了 IBVS 算法，这意味着研究真正的开始。

第3章 伺服目标自主生成算法研究

3.1 引言

第二章完成了 IBVS 的模型建立、仿真系统搭建以及在实物上成功运行 IBVS 算法，这些为后续的研究提供了控制基础。现在机器人知道自己怎么动了，那么它面对未知物体抓取任务时如何知道自己往哪动呢？接下来，在本章中将对本次课题又一基石——伺服目标生成的研究进行展开。主要研究了基于模型的点云识别和抓取点生成网络两种方法，将它们进行比较后，选择了后者作为应对未知物体抓取任务时生成机器人伺服目标的方案。

3.2 基于模型的点云识别与配准

3.2.1 方法陈述

虽然物体样式多种多样，但它们总以一类一类形式呈现，例如不同种类的苹果，形状会类似，苹果和球形状类似。对于每一类这样的物体称为一个模型，而对于每一个模型会有一个确定的抓取方式。先收集尽可能多的点云，计算它们的点云特征，对特征类似的进行聚类，计算每一类特征的平均值作为一个模型。每个模型会人为的设定点云中的部分特别位置的点作为期望特征。制作的其中一个圆柱类模型如图?所示。

对于新来的未知物体，对每个模型进行 ICP 配准，解算当前物体相对模型的姿态，然后与原模型一致的方式提取当前点云特征，匹配模型的期望特征。将期望特征和提取的特征输入到 IBVS 控制器中完成伺服控制。

ICP 即为迭代最近点法。假设有 PA 和 PB 两个点云，它们是相同物体的不同位姿时深度传感器获取的点云。该算法通过不断迭代它们之间的坐标变换矩阵参数的方式让两个点云尽可能的重叠^[3]。设 PA 和 PB 的点云分布分别为 $p(x, y, z)$ 、 $q(x, y, z)$ ，目前需要找到 P 到 Q 的旋转矩阵 R 和平移矩阵 T ，给出一个代价函数，通过最小二乘法求解最优解。在这之前，先表示出两点云的质心：

$$\begin{cases} \vec{p} = \frac{1}{n} \sum_{i=1}^N \vec{p}_i \\ \vec{q} = \frac{1}{n} \sum_{i=1}^N \vec{q}_i \end{cases} \quad (3-1)$$

然后从两个点集中的每个点减去相应的质心：

$$\begin{cases} \vec{p}_t = \vec{p}_i - \vec{p} \\ \vec{q}_t = \vec{q}_i - \vec{q} \end{cases} \quad (3-2)$$

则上述最优化目标函数可以转化为：

$$E = \sum_{i=1}^N |\vec{q}_t - R\vec{p}_t| \quad (3-3)$$

最优化问题最后分解为：

- (1) 求使代价函数 E 最小的旋转矩阵 R 。
- (2) 求得平移矩阵 $T = \vec{q} - R\vec{p}$ 。

3.2.2 算法实现

深度传感器获得的点云往往不全是目标物体的点云，在姿态匹配前，经过三个方向的点云截断滤波、去离群点、降采样操作，最后提取出目标点云。基于模板匹配点云识别过程中所提取的点云特征选取了 VFH 特征，它是一种全局特征，可以快速计算和匹配。算法的实现效果如图3-1和图3-2所示：

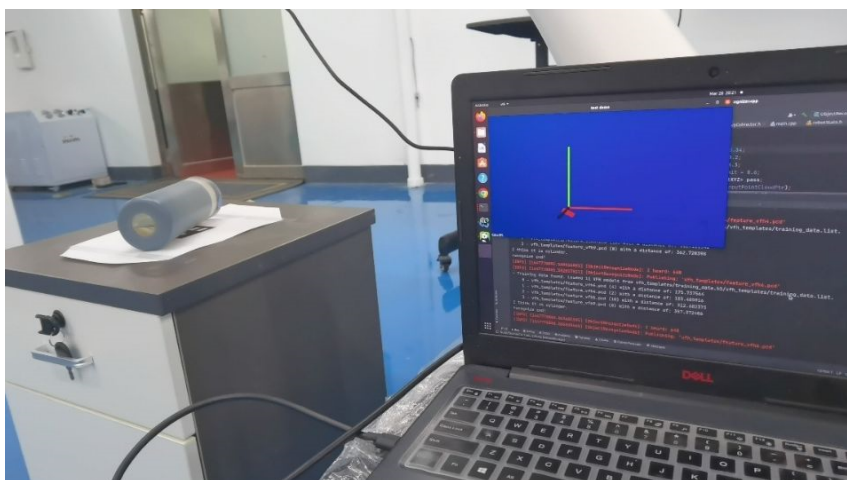


图 3-1 基于模型的点云识别实操图

图3-1和图3-2分别是算法运行时第三方视角和电脑视角的图片，点云处理间隔为 500ms，但识别和位姿解算间隔约为 3s。

3.2.3 方法小结

该方法需要环境噪声较小时，才能正常匹配，消耗时间长，拖长了系统抓取物体所需运行的时间，且不可用于实时的目标特征生成，视觉伺服效果将会非常依赖初始的视觉、点云处理。另外，还存在模型制作困难、鲁棒性差等问题，这写让这个方法面对卷积神经网络完全没有一战之力。

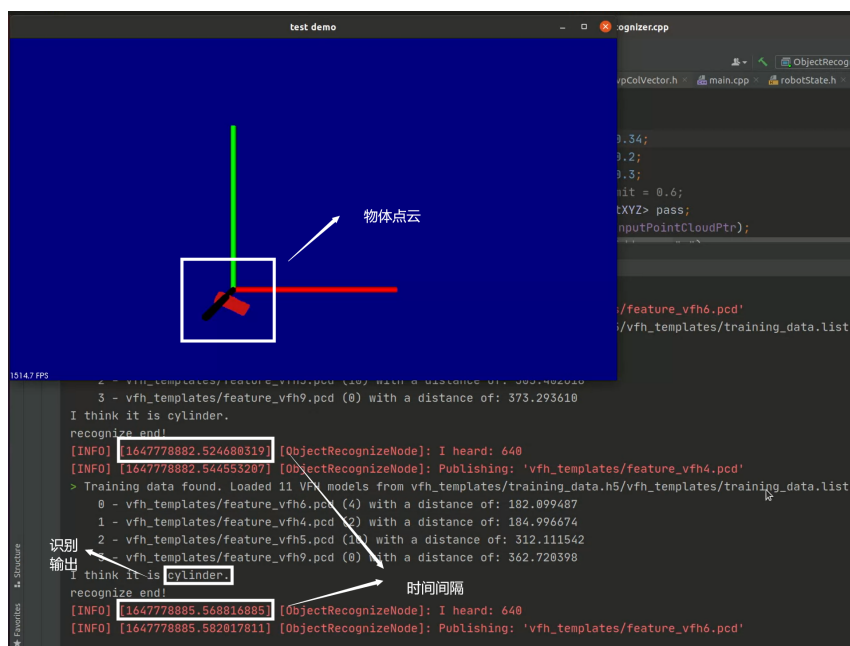


图 3-2 基于模型的点云识别效果图

3.3 抓取点生成网络（GG-CNN）

3.3.1 方法陈述

抓取点生成网络（GG-CNN）属于 Grasp Synthesis 中的一种，为 PBVS 量身订做的抓取点生成方案。Douglas Morrison 等人在 2018 年提出了 GG-CNN，该网络非常轻便快捷，可以通过输入深度图像，在 19ms 内输出图像中机器人的期望抓取位姿，最后依赖相机内参、外参计算出机械臂末端期望的位姿，通过 PBVS 来抓取未知物体^[4]。该工作最大的特点是，它能让视觉伺服实时生成期望位姿，伺服精度不再受初始计算的期望位姿影响。在后续研究中，会通过一个很特别而简单的方法将该网络所输出的结果应用于 IBVS，但在这之前，先要弄清楚它的工作原理。

他们创新性在于提出了十分合适的网络输出。网络整体结构的设计是非常简单的，追求大感受野，然后就是很寻常的叠层。网络限定了机器人末端需要时二指的，视觉伺服控制中，相机必须保持时刻俯视，这也是在第二章实物搭建中这么做的原因之一。网络的本质是语义分割，输出 4 张与输入的 300×300 深度图像相同大小的图像 $G = [Q, W, \Phi(\sin \theta, \cos \theta)]$ ，其中使用 $\sin \theta$ 和 $\cos \theta$ 分别对应不同的输出图片。 Q 图像中每个像素代表这个点的抓取质量，它们都是被归一化的数据，1 表示抓取质量很高，0 表示这个点完全不值得抓取； W 图像中每个像素代表抓取这个点所需要的二指张开宽度； Φ 图像中每个像素代表抓取这个点所需要的二指沿 Z 轴旋转角度。

本文中认为他们所构建的网络有很大的优化空间。他们为了加快网络计算速度，所设计的层数太少，网络的非线性程度较低。所以使用 1×1 卷积层对网络非线性化。其次，作者非常喜欢使用大卷积核。实际上在机械臂末端运动过程中，物体在相机中的大小会有很大的改变，单一感受野并不能适应这样的变化，所以将大卷积核拆成不同尺度小卷积核的叠加。以上的优化并不会带来太多的计算量，因为改动的地方只是增加了 1×1 卷积层和拓展的小卷积核，它们本身不会带来什么计算量。最终网络如图3-3所示：

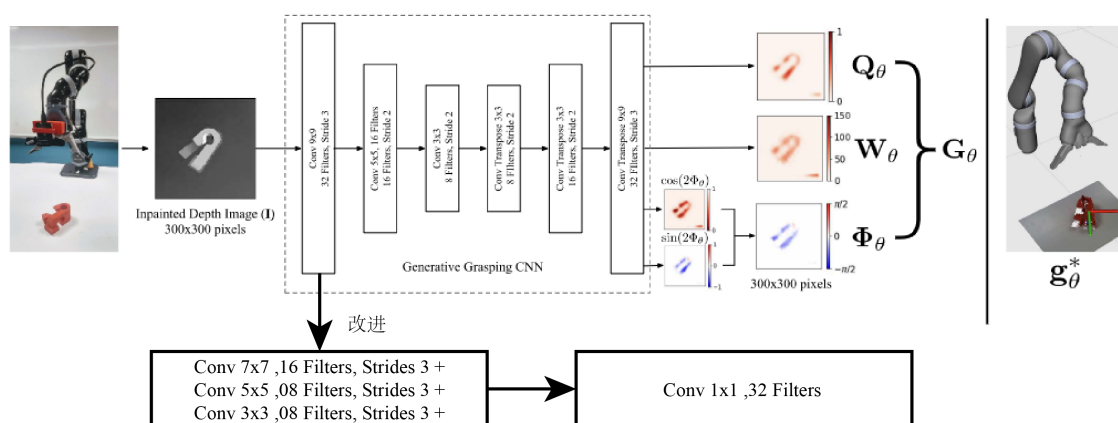


图 3-3 GG-CNN 改进后网络架构

3.3.2 算法实现

配好 pytorch 环境，使用 Cornell 数据集。将计算出的最大抓取质量像素点作为长方形中心， Φ 为长方形绕中心旋转角度， W 作为长方形的长，长方形宽为它的 $1/2$ ，绘制长方形框，将计算出的 IoU（预测抓取框与标签抓取框对应交集与并集的比）作为预测准确率，选取数据的 10% 作为测试集，跑通优化后的 GG-CNN 代码训练程序，约 40 epoch 时达到了对测试集的 80% 准确率。这个准确率相对于原始程序上升了 4%，优化是有效的。对给定深度图，输出效果如图3-4所示：

这里特别解释一下图3-4中图 Q 表示抓取质量热度图，颜色越偏暖色越值得被抓，图 Angle 表示抓取角度图，单位是弧度，可以看到对柱形物体，抓取角度在高抓取质量点处几乎一致。

在代码测试过程中发现 Realsense D435i 对深度的测量信噪比很低，测量深度波动很大，只依靠深度图像一阶图像矩无法稳定定位目标物体的位置，所以通过 RGB 图像和深度图像的一阶图像矩（如果能有更好的深度传感器，是不需要依赖颜色信息的）可以得到目标大致位置。使用这个作为中心对当前 640×480

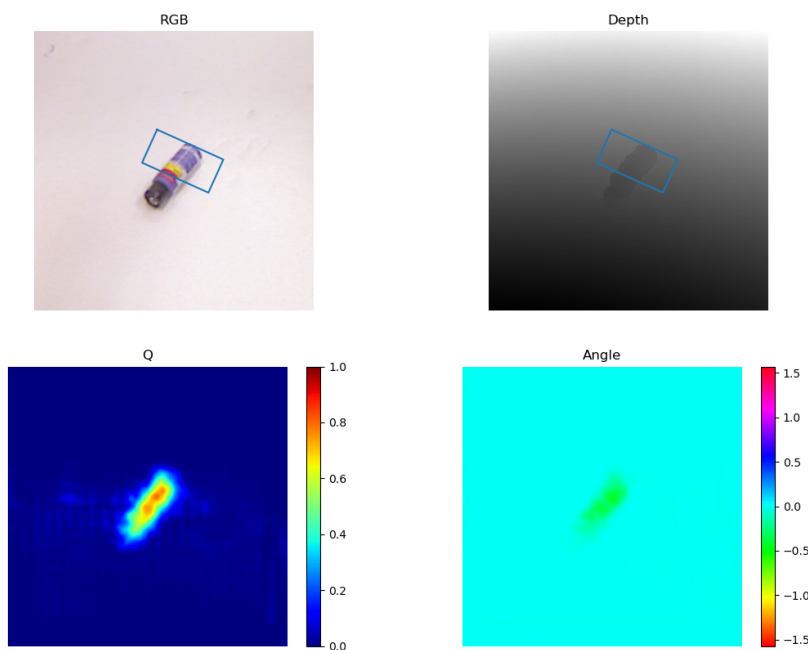


图 3-4 GG-CNN 输出

的图像进行考虑边缘（如果超出原始图像范围，会平移中心）的 300×300 裁剪。将神经网络写成 ROS2 的一个节点，实时发布计算出的抓取点信息，视觉伺服节点会订阅它的主题，实时显示抓取方框。

实时显示抓取框和之前只显示一次抓取框有很大的不同，因为神经网络输出的抓取质量在目标位置处会有很多相近的点，如果物体具有平移对称性或者旋转对称性，抓取点的位置会不断跳动，正如图? 所示。这会导致生成的目标特征不断摆动，导致系统失稳。

在 GG-CNN 原文中给出的解决方式是计算抓取质量图中三个局部最大点，选择与上一次的抓取点最近的点作为本次的抓取点。但是研究中在复现他们的算法后依然无法解决波动很大的问题，甚至调高选择的局部最大点个数到 10。本文认为这是物体平移、旋转对称性带来的必然结果，无论网络好坏，因为网络的输入是深度图像。所以该网络输出的抓取点位置信息无法使用，只能委曲求全使用 RGB 图像的一阶图像矩代替，毕竟在裁剪图像的时候就求取过一阶图像矩的值。颜色信息在所搭建的实验环境中是非常稳定的，这样稳定的抓取点中心可以被用于实时抓取点生成。只是使用该方法会带来另一个麻烦，在伺服末期，如果目标靠近平台边缘，摄像头将捕捉地面的颜色信息，会对抓取点中心的计算带来干扰，这个问题将在本文第四章和其它问题一并解决。

好在角度的波动能得到很好的解决。因为相机时刻俯视向下，所以相机期望的沿 Z 轴的旋转角度和抓取框旋转角度是一致的。对于这样一个线性系统，

使用卡尔曼滤波可以有效抑制噪声。机械臂末端沿 Z 轴的角速度设为 ω_z ，它由每次控制周期 δ_t 的 IBVS 控制器计算出。机械臂末端旋转角度为 θ_z ，可以通过读取机械臂编码器的值间接获取。可以写出状态、观测方程：

$$\hat{x}_{k+1} = \begin{bmatrix} \theta_z \\ \omega_z \end{bmatrix}_{k+1} = \begin{bmatrix} 1 & 0 \\ 0 & \delta_t \end{bmatrix} \begin{bmatrix} \theta_z \\ \omega_z \end{bmatrix}_k = A\hat{x}_k \quad (3-4)$$

$$z_{k+1} = \theta_z = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \theta_z \\ \omega_z \end{bmatrix}_k = H\hat{x}_k \quad (3-5)$$

认为过程噪声 ω_k 和观测噪声 v_k 都服从高斯分布：

$$\begin{cases} p(\omega) \sim N(0, Q) \\ p(v) \sim N(0, R) \end{cases} \quad (3-6)$$

则有协方差矩阵的预测方程：

$$\hat{P}_{k+1} = A P_k A^T + Q \quad (3-7)$$

设卡尔曼滤波的增益为 K ，则根据以上条件求得：

$$K_{k+1} = \hat{P}_{k+1} H^T (H \hat{P}_{k+1} H^T + R)^{-1} \quad (3-8)$$

使用观测器数据对当前预测状态进行更新，对卡尔曼滤波器的初始化中，令 R 、 Q 都为单位矩阵：

$$x_{k+1} = \hat{x}_{k+1} + K_{k+1} (z_{k+1} - H\hat{x}_{k+1}) \quad (3-9)$$

再使用新的预测的状态更新协方差矩阵：

$$P_{k+1} = (I - K_{k+1}H) \hat{P}_{k+1} \quad (3-10)$$

在机械臂静止不动时，神经网络预测的抓取点中心、角度随时间变化曲线图? 所示。得到的抓取中心只有 1 到 2 像素点的波动，预测的抓取角度波动几乎消失了多少?。

3.3.3 方法小结

GG-CNN 是十分契合本次研究目的的研究成果，因此它被选作未知物体抓取算法中解决机械臂怎么去抓取目标的方法。但是在复现工作中遇到了抓取中心波动过大的无法解决的问题，只能使用 RGB 图像的一阶图像矩方法代替它

进行抓取点中心生成，最终只保留了该网络输出的角度项，在经过卡尔曼滤波处理后拥有比较稳定的输出值。

3.4 本章小结

本章研究了两种伺服目标自主生成的算法，比较后，选择了后者作为后续研究生成伺服目标的主要方法。相比于传统的基于模型获取抓取点的方法，神经网络有更好的鲁棒性、实时性。但是目前选取的 GG-CNN 并不是完美的，在实时伺服中它的抓取点中心因为目标物体平移、旋转对称性的存在，波动已经大到无法使用滤波的方法来抑制了。所以使用 RGB 一阶图像矩来作为抓取点中心的生成依据。对于网络输出的抓取角度使用线性卡尔曼滤波对波动进行有效的抑制。

第 4 章 未知物体抓取算法实现

4.1 引言

4.2 机器人视觉伺服系统概述

第 5 章 IBVS 控制律优化

第 6 章 实验设计与验证

6.1 引言

前面的章节中详细说明了通过视觉伺服抓取未知物体的思路、视觉伺服控制律的优化以及它们的实现过程。在前人的研究成果基础上，这套未知物体抓取算法提出创新性的特征点提取、匹配方式和与之配套的视觉伺服控制律。它将在更少的环境条件限制下拥有更高的未知物体抓取率和更理想平滑的曲线。本章将对一共八个家常物体，每个物体八种位姿进行抓取实验，然后将抓取成功率与各生成抓取合成类方法进行比较，证明这套未知物体抓取算法的优越性。另外，会对不同物体同一位姿、相同物体不同位姿的伺服曲线进行记录，通过超调、响应时间等参数反映算法应对多种多样情况下的高鲁棒性。

6.2 实验设计

对于未知物体抓取任务，抓取成功率是最需要关注的指标。为了使成功率的测试结果能体现系统应对各种情形下的抓取能力，需要实验中的抓取目标种类多样化和位姿多样化。另外，为了能与提出 GG-CNN 原文献抓取成功率结果相对公平性的横向比较，尽量搭建相同的实验环境和使用相同的抓取目标。

根据算法的实现过程，实验过程中存在的限制条件如下：

(1) 由于研究中 GG-CNN 对输入深度图像的尺寸需求和清晰度需求，需要对目标的进行范围裁剪，这时会使用 RGB 信息确定目标的中心大致范围，所以实验中在平台上铺盖的白纸是必须的。如果有更好的网络能不需要裁剪并稳定指出目标位置，实验环境中白色背景的条件可以被去除。

(2) GG-CNN 的输出形式决定了如果要使用它的输出结果作为视觉伺服的依据，必须保证相机时刻俯视 0° 朝下，所以对末端的沿 X、Y 轴角速度指令不论计算结果如何都会在给机械臂下达命令那一步置零。

(3) 深度传感器使用的 Realsense D435i，它是基于结构光的原理测量深度的，所以对漫反射能力差的目标，获取的深度值几乎是无效值，所以选用的目标都是漫反射效果好的物体。另外，它的深度测量精度为 1mm，这决定了选用目标不可以是铅笔等过小的物体，当然，也不可以过大，这会让神经网络无法区分谁是目标谁是背景。

实物环境搭建如图2-8，这与 GG-CNN 原文是类似的。但值得一提的是，经过第四章中对 RGB 图像和深度图像的特别处理，本文中机械臂的运动范围可以超出平台范围（但初始情况时，目标必须在相机视野范围内），这是相对 GG-CNN 原文具有优势的地方。根据限制条件（3）和仿照 GG-CNN 原文使用的抓取目标，选择了八种家常抓取目标用于本次实验的各种测试与验证，它们是什么正如图所示：

抓取目标将以对称的姿态在平台的八个方位的位置摆放，这么做可以很全面地测试伺服系统应对多样化的目标位姿的伺服情况，摆放示意图如图? 所示：

六轴 JK 机械臂初始状态各编码器显示数值为：机械臂末端在 {B} 中的初始状态为:。在这样的初始参数设计下，摄像头获取的图片如图? 所示，正好可以将整个平台尽收眼底。相对机械臂末端，目标摆放位置 and 对应姿态如表? 所示

实验中将会在八种不同的物体的八种不同位姿情况下对目标进行抓取，记录的实验结果包括是否成功抓取，伺服超调量和响应时间。响应时间不包括抓物体所需时间，因为抓取是开环控制的，所需时间固定，没有记录的必要。开环抓取效果在第四章有详细介绍。然后会与各个基于生成抓取合成方法来抓取未知物体的文献成果进行比较。最后对不同物体相同位姿以及抓取率百分之百的一个物体不同位姿伺服曲线进行分析，并与同样基于 GG-CNN 但使用 PBVS+IBVS 来抓取未知物体的文献成果^[5] 进行比较。

6.3 人机交互界面设计

6.4 实验验证

6.4.1 抓取成功率测试

对总共 64 种情况的实验结果进行记录，不仅记录总的抓取成功率，同时记录相同物体的抓取成功率和相同位姿对应不同物体的抓取成功率，绘制表格如下：

通过该表格展现的数据，反应所提出的未知物体抓取算法特点如下：

将实验获得的数据与各类成抓取合成方法的文献成果在一张表格中进行展现：

表格反映了。。。

6.4.2 抓取性能验证

选择的目标为。。。, 把它不同的摆放位姿情况下的特征偏差曲线绘制到同一坐标系下, 将不同情况对应的超调量和响应时间绘制到直方图中, 如图? 所示:

从图? 中可以看出。。。

选择的位姿为。。。, 把对应的不同目标情况下的特征偏差曲线绘制到同一坐标系下, 将不同情况对应的超调量和响应时间绘制到直方图中, 如图? 所示:

图? 为同样基于 GG-CNN 但使用 PBVS+IBVS 来抓取未知物体的伺服过程中特征偏差曲线, 他还测量了运动物体的伺服情况, 相关曲线忽略。比较静止物体的伺服情况, 有以下结论。。。

6.5 本章小结

本章在八种不同家常的物体的八种不同位姿情况下对目标进行抓取实验, 在所提出的特征提取、匹配的方法和与之配套的伺服控制律的加持下, 在环境限制条件更少的情况下, 拥有比许多类似工作的研究成果更高的抓取成功率。与另一个基于 GG-CNN 抓取未知目标文献成果比较, 拥有更平滑而理想的伺服曲线。通过以上实验证明了本研究所提出的未知物体抓取算法有高鲁棒性和高伺服性能。

结 论

学位论文的结论作为论文正文的最后一章单独排写，但不加章标题序号。

结论应是作者在学位论文研究过程中所取得的创新性成果的概要总结，不能与摘要混为一谈。博士学位论文结论应包括论文的主要结果、创新点、展望三部分，在结论中应概括论文的核心观点，明确、客观地指出本研究内容的创新性成果（含新见解、新观点、方法创新、技术创新、理论创新），并指出今后进一步在本研究方向进行研究工作的展望与设想。对所取得的创新性成果应注意从定性和定量两方面给出科学、准确的评价，分（1）、（2）、（3）…条列出，宜用“提出了”、“建立了”等词叙述。

参考文献

- [1] 孙冬雪. 基于模型预测方法的机器人视觉伺服控制研究[D]. 长春: 长春工业大学, 2018: 9-10.
- [2] Chaumette F, Hutchinson S. Visual servo control. I. Basic approaches[J]. IEEE Robotics & Automation Magazine, 2006, 13(4): 82-90.
- [3] Chetverikov D, Svirko D, Stepanov D, et al. The trimmed iterative closest point algorithm[C] // Object recognition supported by user interaction for service robots: Vol 3. 2002: 545-548.
- [4] Morrison D, Corke P, Leitner J. Closing the loop for robotic grasping: A real-time, generative grasp synthesis approach[J]. arXiv preprint arXiv:1804.05172, 2018.
- [5] Haviland J, Dayoub F, Corke P. Control of the final-phase of closed-loop visual grasping using image-based visual servoing[J]. arXiv preprint arXiv:2001.05650, 2020.

哈尔滨工业大学本科毕业设计（论文）原创性声明

本人郑重声明：在哈尔滨工业大学攻读学士学位期间，所提交的毕业设计（论文）《基于抓取点生成网络与视觉伺服的未知物体抓取算法研究》，是本人在导师指导下独立进行研究工作所取得的成果。对本文的研究工作做出重要贡献的个人和集体，均已在文中以明确方式注明，其它未注明部分不包含他人已发表或撰写过的研究成果，不存在购买、由他人代写、剽窃和伪造数据等作假行为。

本人愿为此声明承担法律责任。

作者签名：

日期： 年 月 日

致 谢

衷心感谢导师 XXX 教授对本人的精心指导。他的言传身教将使我终生受益。

.....

感谢哈工大 L^AT_EX 论文模板 hiThesis !