

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

毕业设计（论文）

题 目 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

专 业 自动化

学 号 1181140118

学 生 王泽飞

指 导 教 师 高会军

答 辩 日 期 2022 年 6 月 x 日

摘要

在工业、服务业智能化发展的大背景下，视觉伺服几乎已是现在自动化工厂不可缺少的部分。由于产品寿命短，生产线经常变动，工业机器人不可避免的要从固定模式的运动控制到更为灵活的行为模式中去。

本文将针对未知物体的 IBVS 抓取算法中涉及到的抓取点自主生成、图像特征的稳定提取与匹配、IBVS 伺服性能提升等关键问题进行深入研究。

主要研究内容包括以下几个方面：

首先，建立了 IBVS 系统模型，其中包括系统坐标系和视觉坐标系的建立，推导了各坐标系之间的变换关系以及 IBVS 要求的特征点位置获取的公式。研究了基于特征点交互矩阵的 IBVS 的原理，并搭建仿真环境。搭建了实物运行环境，成功运行 IBVS 算法。

其次，针对伺服目标如何自主生成的问题研究了基于模型的点云识别和配准以及抓取点生成网络两种方式的原理和生成效果。根据它们的实时性、可靠性确定了抓取点生成网络和一阶图像矩搭配的方式作为伺服目标自主生成的方案。优化了抓取点生成网络提升了网络预测正确率。使用卡尔曼滤波算法让网络预测的角度输出变得稳定。

再次，结合前面研究的内容对未知形状、位姿随机的目标的抓取展开研究。结合 ORB 特征的提取与匹配，提出了一种基于抓取点的 IBVS，根据不断失败的经验，又提出了基于抓取点的自定义特征。该方法解决了速度指令波动大和关于特征点匹配的可靠性、顺序一致性等问题。面对当前伺服性能差的问题，对当前特征偏差曲线进行分析，优化了 IBVS 控制律。对于出现的特征偏差对末端速度指令关联的不对称性问题和噪声问题提出一种自适应的算法进行解决。

最后，设计实验，通过多样的物体、多样的位姿的抓取实验验证提出的未知物体抓取算法的有效性。与相关工作进行比较，证明当前算法的高自适应能力和高伺服性能。

关键词：机器人；视觉伺服；生成抓取合成；目标抓取

Abstract

In the context of the intelligent development of industry and service industries, visual servoing has almost become an indispensable part of automated factories. Due to the short product life and frequent changes of production lines, it is inevitable for industrial robots to move from a fixed mode of motion control to a more flexible behavioral mode.

This paper will conduct in-depth research on the key issues involved in the IBVS grasping algorithm of unknown objects, such as the autonomous generation of grasping points, the stable extraction and matching of image features, and the improvement of IBVS servo performance.

The main research contents include the following aspects:

First, the IBVS system model is established, including the establishment of the system coordinate system and the visual coordinate system, and the transformation relationship between the coordinate systems and the formula for obtaining the position of the feature points required by IBVS are deduced. The principle of IBVS based on feature point interaction matrix is studied, and the simulation environment is built. A physical operating environment was built, and the IBVS algorithm was successfully run.

Secondly, the principle and generation effect of two methods of model-based point cloud recognition and registration and grasp point generation network are studied for the problem of how to generate servo targets autonomously. According to their real-time and reliability, the method of generating network of grab points and first-order image moments is determined as the scheme for autonomous generation of servo targets. Optimized the grab point generation network to improve the accuracy of network prediction. The Kalman filter algorithm is used to stabilize the angle output predicted by the network.

Thirdly, combined with the content of the previous research, the grasp of the unknown shape and random pose is studied. Combined with the extraction and matching of ORB features, an IBVS based on grasp points is proposed. According to the experience of continuous failure, a custom feature based on grasp points is proposed. This method solves the problems of large fluctuation of speed command, reliability and order consistency of feature point matching. Facing the problem of poor current servo performance, the current characteristic deviation was analyzed in time domain and frequency domain, and the IBVS

control law was optimized. An adaptive algorithm is proposed to solve the asymmetry and noise problems associated with the terminal speed command due to the characteristic deviation.

Finally, experiments are designed to verify the effectiveness of the proposed unknown object grasping algorithm through grasping experiments of various objects and various poses. The comparison with related work demonstrates the high adaptive ability and high servo performance of the current algorithm.

Keywords: Robot, Visual servo, Generative Grasp Synthesis, Target graph

目 录

摘要	I
Abstract	II
第1章 绪论	1
1.1 课题背景及研究的目的和意义	1
1.2 机器人视觉伺服发展现状	2
1.3 基于视觉伺服的物体抓取方法概述	4
1.3.1 抓取方法	4
1.3.2 伺服性能	5
1.4 主要研究内容及章节安排	5
第2章 基于图像的视觉伺服（IBVS）理论研究	7
2.1 引言	7
2.2 IBVS 系统模型建立	7
2.2.1 系统坐标系建立	7
2.2.2 视觉模型建立	8
2.3 IBVS 仿真系统实现	9
2.3.1 IBVS 算法原理	9
2.3.2 机器人仿真模型搭建	10
2.3.3 曲线绘制与相机轨迹记录	11
2.4 IBVS 实际系统实现	12
2.4.1 IBVS 实物系统平台搭建	12
2.4.2 IBVS 实际运行	13
2.5 本章小结	14
第3章 伺服目标生成算法研究	15
3.1 引言	15
3.2 基于模型的点云识别与配准	15
3.2.1 方法陈述	15
3.2.2 算法实现	16
3.2.3 方法小结	17

3.3 抓取点生成网络（GG-CNN）	18
3.3.1 方法陈述	18
3.3.2 算法实现	18
3.3.3 方法小结	22
3.4 本章小结.....	22
第4章 未知物体抓取算法实现	23
4.1 引言	23
4.2 图像处理抑制边界环境干扰	23
4.3 初步思路.....	25
4.3.1 ORB 特征的提取与匹配.....	25
4.3.2 基于抓取点的 IBVS	27
4.3.3 总结与思考	29
4.4 一种基于抓取点的自定义特征	29
4.4.1 方法概述	29
4.4.2 目标表面计算	30
4.4.3 优化特征分布与 IBVS 试验.....	32
4.5 本章小结.....	33
第5章 IBVS 控制律优化	34
5.1 引言	34
5.2 控制模型建立	34
5.3 PD 控制器与惯性环节	34
5.4 滑模控制器	34
5.5 系统稳定性	34
5.6 特征偏差关联速度不一致性	34
5.7 本章小结.....	34
第6章 实验设计与验证	35
6.1 引言	35
6.2 实验设计.....	35
6.3 实验验证.....	36
6.3.1 抓取成功率测试	36
6.3.2 抓取性能验证	37
6.4 本章小结.....	37
结 论	39

参考文献	41
哈尔滨工业大学本科毕业设计（论文）原创性声明	43
致 谢	44

第1章 绪论

1.1 课题背景及研究的目的和意义

随着工业、服务业智能化发展，机器人需要在更苛刻的环境下自主地完成更严格要求的操作。视觉伺服的不接触性让智能抓取技术变得灵活可靠，硬件的制造工艺快速进步使其稳定性、实时性得到很大程度的增长。而相比传统机器人，采用视觉伺服的方式不仅能让处理机器人控制问题时拥有更充分直接的信息，而且其非接触式的特点让机器人能更灵活地工作^[1,2]。

通常情况下，我们使用教学的方式让机械臂记录如何去完成自己的抓取动作，因为给予它们的任务往往是简单而重复的。然而在出现抓取目标多样化，甚至只是出现位姿的改变，都需要重新进行教学^[3]。另外，这样的系统对机器人驱动误差或传感器误差没有任何鲁棒性，机器人损坏后更换完各个设备，又将面临重新调参，这完全是可以避免的。人工控制更为灵活，但人需要长时间的培养和经验积累才能对机器较为精准快速地控制，并且直接控制的精准和稳定性已经无法满足现代工厂的需求了。机器人需要能自主根据环境进行决策。自主可靠的解决方案中很难不考虑视觉伺服。视觉伺服几乎已是现在自动化工厂不可缺少的部分。由于产品寿命短，生产线经常变动，工业机器人不可避免的要从固定模式的运动控制到更为灵活的行为模式中去。它能让机器人具有“眼睛”这样一个强力的感官，视觉信息更能适应于大部分场景，以应对不同的任务需求。这么做不但减少人力，节约了劳动力资源，而且产品的质量、生产效率都有很大程度的提升^[4]。但当今视觉伺服在工业生产的应用中仍存在自适应能力差、精度低、稳定性不够等技术问题。



图 1-1 工业机器人与视觉伺服



图 1-2 医用机器人与视觉伺服

智能服务型机器人是当今另一个机器人领域中很火热的方向，医疗服务是服务型机器人一大应用。外科手术机器人在多年前已被广泛投入使用，医生可以通过摄像头采集到的视觉信息，自身在另一端遥控患者这一端的机械臂操纵手术用具对患者完成手术，这不但确保了手术的精细程度，而且降低了患者手术风险及感染机率^[5]。但这样的技术还是开环控制，更多的依然依赖大夫本身通过视觉图像的判断，人为主动因素仍较高^[6,7]。对于现在的微创手术，精细程度要求更高需要机器人更主动精准的定位和控制。而视觉伺服引入医疗服务，机器人能自动通过图像处理得到患者病灶区精准位置，再通过手术导航软件的引导，手术机械能自动运动到病灶区域，补偿人为控制的误差进行手术^[8]，这种技术能更进一步提高手术精度。如果视觉伺服的自适应定位性能和实时性不够，这些好的思想无法应用于实际，医疗机器人也难以在这个领域发展。

综上所述，研究自主可靠的机器人视觉伺服系统至关重要。本文将针对未知物体抓取算法展开研究，这包括能够在非结构化环境中运行，同时抓住以前从未见过的物品。对于一个更自动化的机器人视觉伺服系统，它持有更少的先验信息（未知物体、随意位姿）但也能很好的完成抓取任务。这是在机械臂视觉伺服抓取领域中大家都在攻克的难关。因为这项功能的完美实现可以大幅度降低工业生产、物流运输等机械臂的调试、维修成本，减少人力操作和节约劳动力资源。所以对机械臂抓取未知物体算法的研究十分具有意义。

1.2 机器人视觉伺服发展现状

常见的视觉伺服分基于图像的视觉伺服（IBVS）、基于位置的视觉伺服（PBVS）和混合视觉伺服（HVS）。IBVS 将图像中的特征相对期望特征的偏差作为控制系统的输入控制执行器速度直至相机中的特征到达期望特征为止。该方法对深度测量具有很好的鲁棒性，但精度依赖于图像信噪比，且是一种局部收敛的控制，如果这个偏差过大，视觉伺服系统将会发散。PBVS 通过图像中的目标位置计算出世界坐标系中目标的位姿，将当前执行器位姿和目标位姿做差输入控制器中。该方法具有全局收敛的特点，但十分依赖内外参、深度测量的精度，同时不可避免的需要轨迹规划。HVS 结合了上述两种方法的特点，但算法实现复杂、可调节参数较多。

物体抓取任务一般出现在工业生产和物流运输中，运作具有局限性，抓取控制中全局收敛是不必要的。另外，这样的抓取环境往往多样而复杂，使用 PBVS 面对不同的抓取场景都要标定外参，这项工作是繁琐且多余的。所以本文集中

研究基于 IBVS 的未知物体抓取算法。

F Chaumette, S Hutchinson 对 IBVS 的最基础的原理进行了详细说明^[9]。该方法使用的特征是图像中提取的特征点，同时也对应到现实世界中，利用它们的信息计算交互矩阵，从而计算机械臂末端速度指令。所以为了保证特征提取与匹配的稳定性，常常使用人工制作的特征。它们的另一篇文章讲述了期望、当前交互矩阵的平均值作为新的交互矩阵有更好的伺服效果^[10]。图1-3是在复现他们成果时的截图。

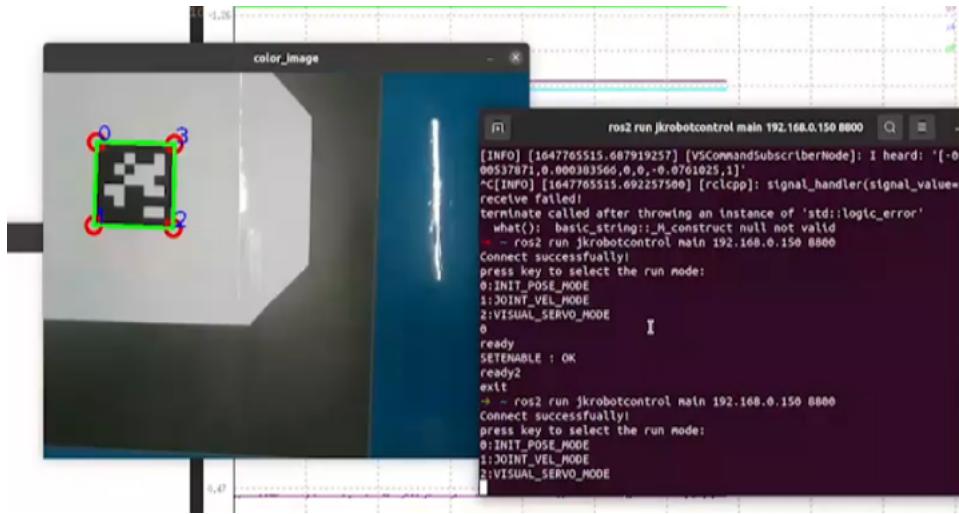


图 1-3 IBVS 基础成果复现

图像矩早在 1960 年被提出并广泛应用于模式识别，它是一种全局特征，旨在克服传统特征提取受噪声影响较大的困难。F. Chaumette 将图像矩用于 IBVS 中，并推导出了基于图像矩的交互矩阵^[11]，它的想法在对平面物体和对称物体的试验中得到很好的实现。传统特征点检测器对特征点检测的不稳定问题可以通过使用图像矩解决，基于图像矩的 IBVS 在后续发展中大放光彩。

C Liu, R Chen 等人提出了基于集空间的 IBVS^[12]，是当时十分新颖的方法，不需要用于图像特征的提取、匹配和跟踪的复杂的图像处理技术。相反，它只需要一个简单的匹配算法并在集合空间中构建视觉偏差。每个错误主要与一个相机的自由度有关，它们设计了一个解耦控制律解决了这个问题。该方法鲁棒性良好，不需要用到相机内参。

随着神经网络的再兴起，在 IBVS 中也卷起了神经网络热潮。F Tokuda, S Arai 等人提出一种端到端的神经网络^[13]，自主提取特征并与目标图像的特征匹配，直接获取末端速度指令。这是一种几乎完全依赖数据的简单粗暴的方式。

综合以上提到的实现 IBVS 的思路，科研工作者几乎都在寻找一种新的特征提取方式来解决传统方法实时性差、兴趣点获取不稳定、鲁棒性差等问题。

方法的发展也逐渐变得自适应，应用范围变得不再局限。

1.3 基于视觉伺服的物体抓取方法概述

基于视觉伺服的物体抓取的核心集中于两个问题：抓取方法和伺服性能。

1.3.1 抓取方法

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。但是这种做法下泛化性是不行的，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。所以为了解决这些问题，多种多样的方法也应运而生。

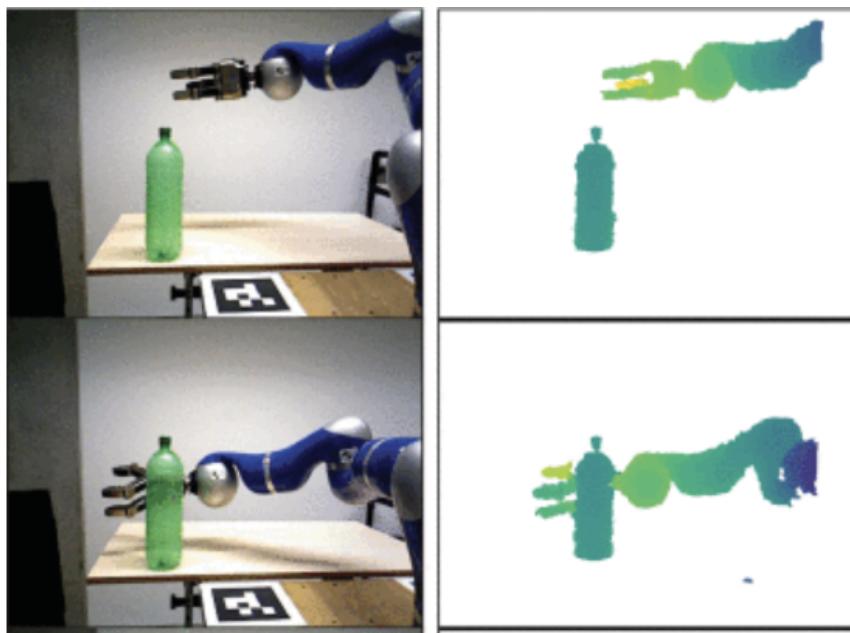


图 1-4 语义分割与 IBVS

发展逐渐成熟的抓取合成技术 (grasp synthesis)，通过抓取图像的图像特征分析出抓取该目标时机械臂应处于的位姿，以此来规划机械臂的伺服过程^[14]。该方法分经验法和深度法，随着深度网络的发展，深度法已经成为抓取合成技术的主要趋势。通过神经网络自主生成目标图像是前几年许多研究者投入的领域之一，但还是相同的问题，这样必须要求目标与当前相似，没有解决伺服范围太小的问题。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^[15]，该方法从一定程度上解决了目标需要与当前相似的问题，但是具有伺服指令不够平滑的缺陷。针对眼到手 (eye-to-hand) 系统的神经网络，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，较为复杂^[16]，图1-4为他们是语义分割的成果图。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。

1.3.2 伺服性能

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，如何结合当前伺服特性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。

1.4 主要研究内容及章节安排

本文以手在眼机械臂视觉伺服系统为研究对象，以抓取未知形状随机位姿的物体为目标展开研究，旨在面对多样的抓取对象，与其它类似方法比较也能拥有更高的抓取成功率。为此，首先搭建IBVS系统模型，在仿真和实物上实现IBVS算法。研究能自主生成抓取期望的算法，选择合适的方法并优化它。对于抓取点生成网络在IBVS中的使用，提出一种新的IBVS特征获取方法，保证特征匹配的准确性。在当前算法基础上给出适合它的IBVS控制律，针对特征偏差对末端速度指令关联的不对称性问题和噪声问题提出一种自适应控制算法。最后通过实物实验证提出的基于抓取点生成网络和IBVS的未知物体抓取算法的高抓取率、高伺服性能。

本文共分为六章，章节内容如下：

第一章：介绍了课题背景及研究的目的和意义，陈述了机器人视觉伺服的发展历程和现状。然后讲述了都有什么方法来实现基于视觉伺服的物体抓取，比较他们的优劣，最终确定自己在这个问题上的研究、实施方案。

第二章：针对未知物体抓取任务搭建IBVS系统模型，其中包括系统坐标系和视觉坐标系的建立。针对以上建立的模型和基于特征点交互矩阵的IBVS原理，实现IBVS系统仿真，并实现数据曲线记录和相机轨迹绘制功能方便调试。最后搭建实物平台，在实物上成功运行IBVS算法。

第三章：研究了基于模型的点云识别与配准和抓取点生成网络(GG-CNN)

两种生成伺服目标的方法，通过比较它们的优劣选择了后者作为伺服目标生成方案。最后采取滤波等方式抑制了神经网络的输出波动。

第四章：通过图像处理解决各种环境干扰问题。将抓取点生成网络运用到IBVS中，提出新的IBVS特征获取方法，实现了未知物体抓取算法并在实物上成功运行。

第五章：针对当前伺服性能差的问题，从时域、频域上给出IBVS控制律的改善方法，相比于原始的伺服指令曲线和特征偏差曲线，改进后的曲线更加理想且光滑。对于出现的特征偏差对末端速度指令关联的不对称性问题和噪声问题提出一种自适应控制算法。

第六章：最后，设计实验，通过多样的物体、多样的位姿的抓取实验验证提出的未知物体抓取算法的有效性。与相关工作进行比较，证明当前算法的高自适应能力和高伺服性能。

章节之间的关系如图1-5所示：

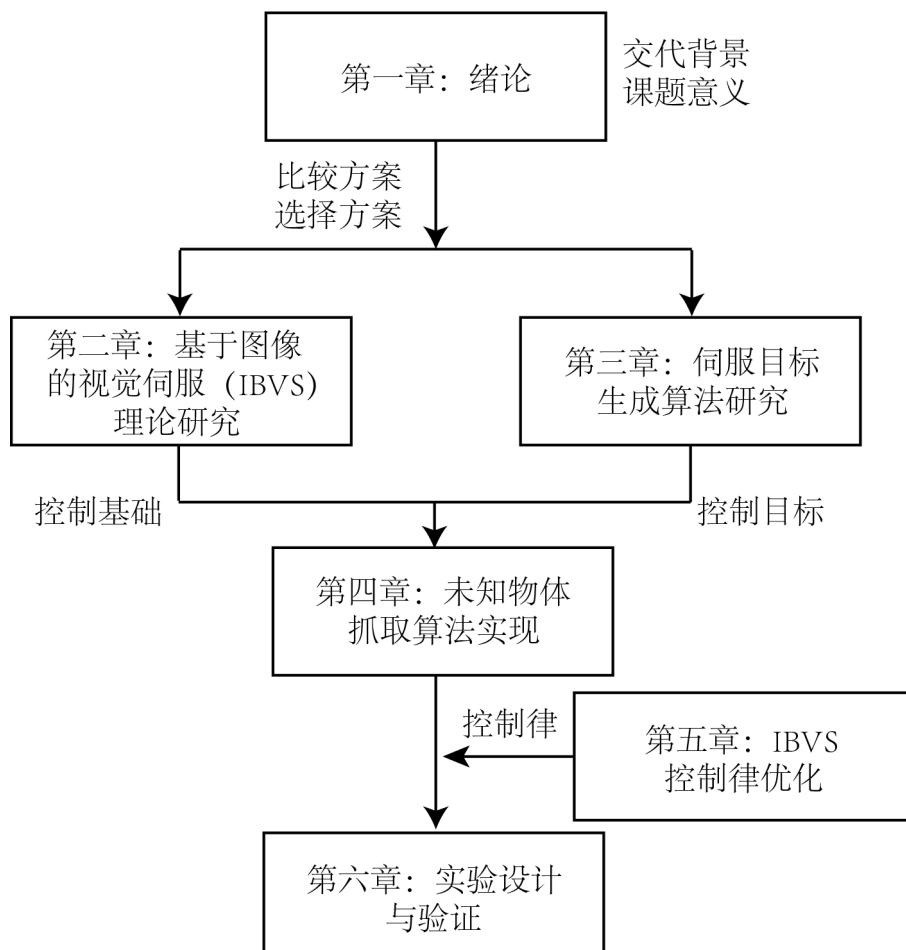


图 1-5 各章节关系图

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：

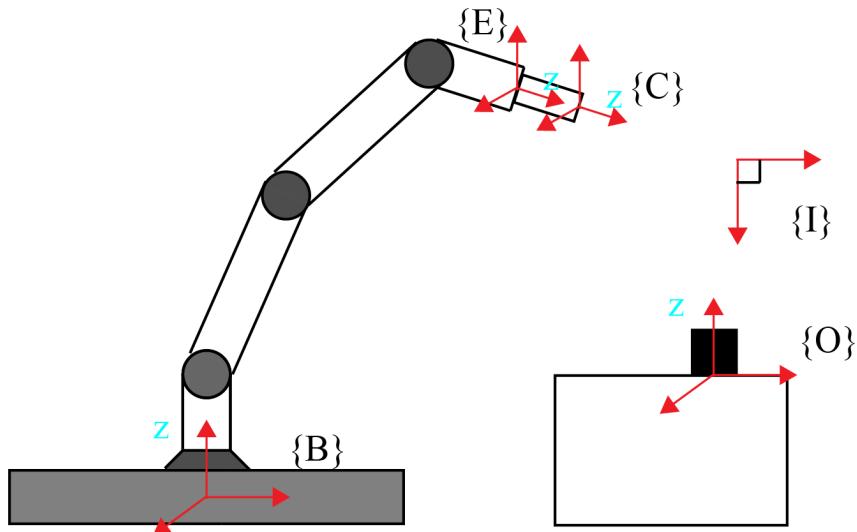


图 2-1 系统坐标系建立示意图

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^[4]。

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：

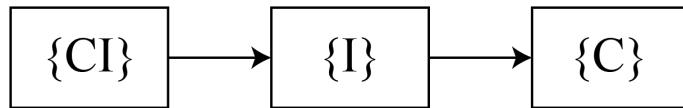


图 2-2 特征变换过程图

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^[4]，这张针孔模型示意图十分典型。

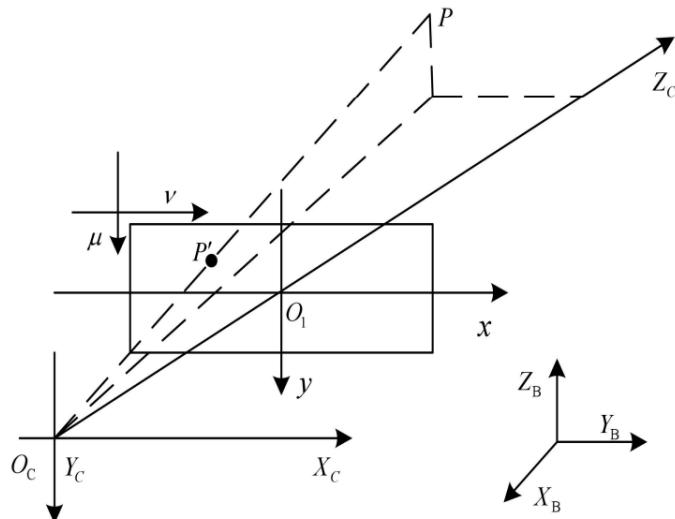


图 2-3 视觉坐标系建立示意图

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^[9]。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^[9], 于是可以得到方程组2-3:

$$\left\{ \begin{array}{l} (\dot{s} - \dot{s}^*) = \dot{e} = L_c v_c \\ L_e = L_c \\ \dot{e} = -\lambda e \end{array} \right. \quad (2-3)$$

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

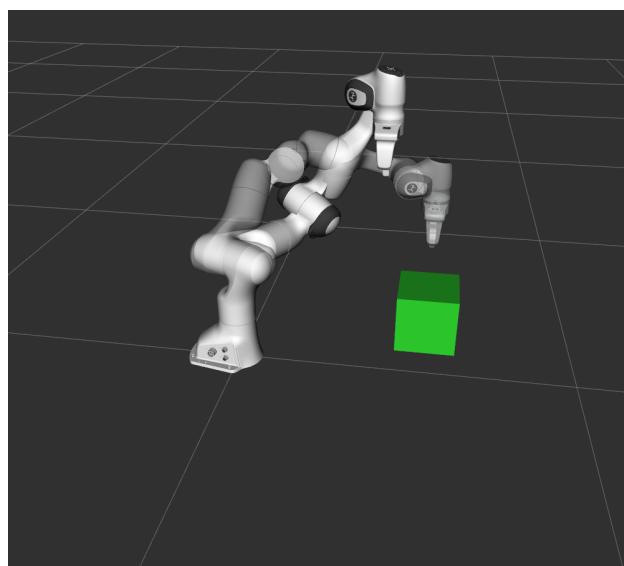


图 2-4 基于 moveit 机器人仿真模型实现

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。

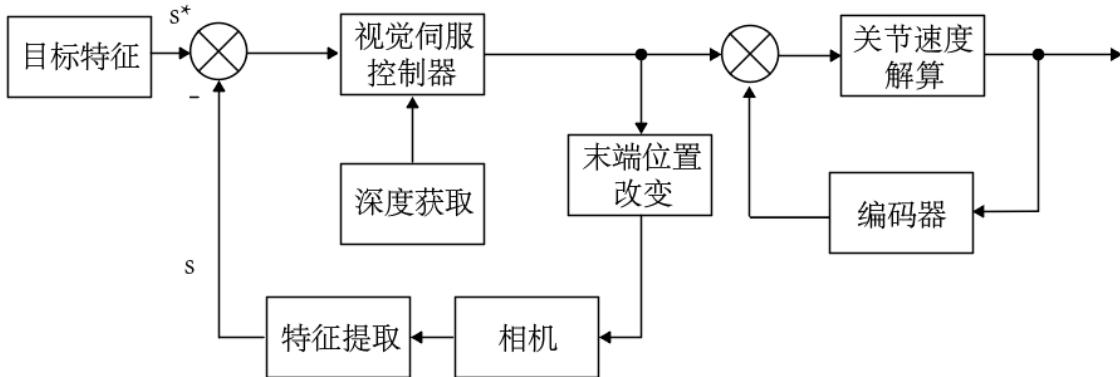


图 2-5 IBVS 基础控制流程方框图

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。

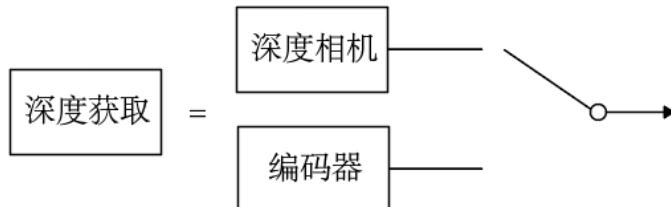


图 2-6 深度传感器切换示意图

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。

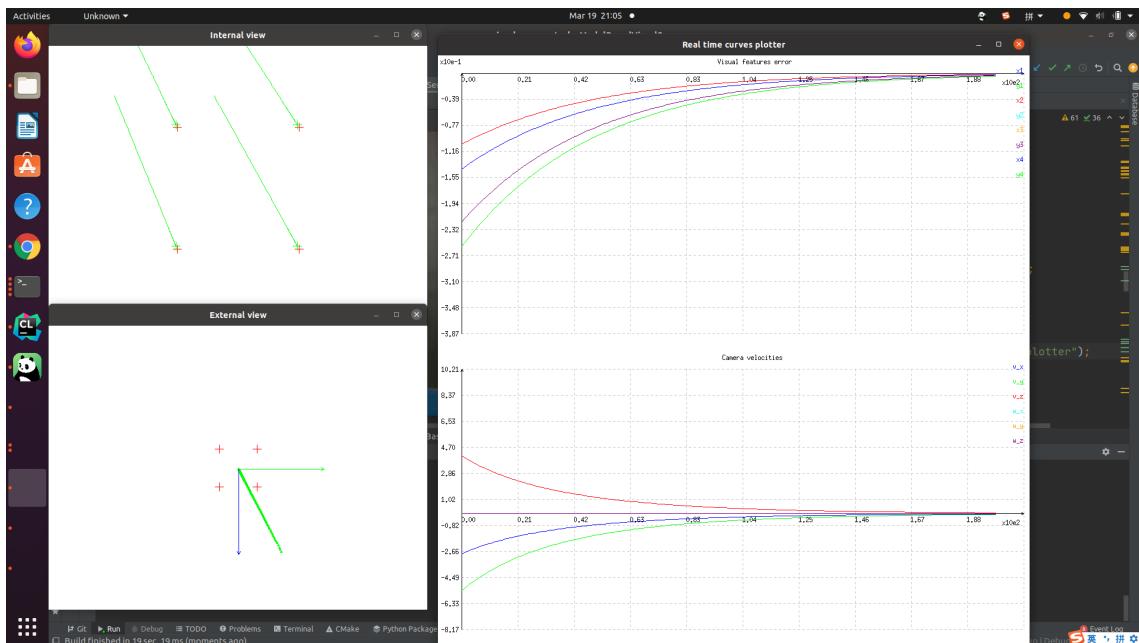


图 2-7 基于 VISP 视觉伺服仿真

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：

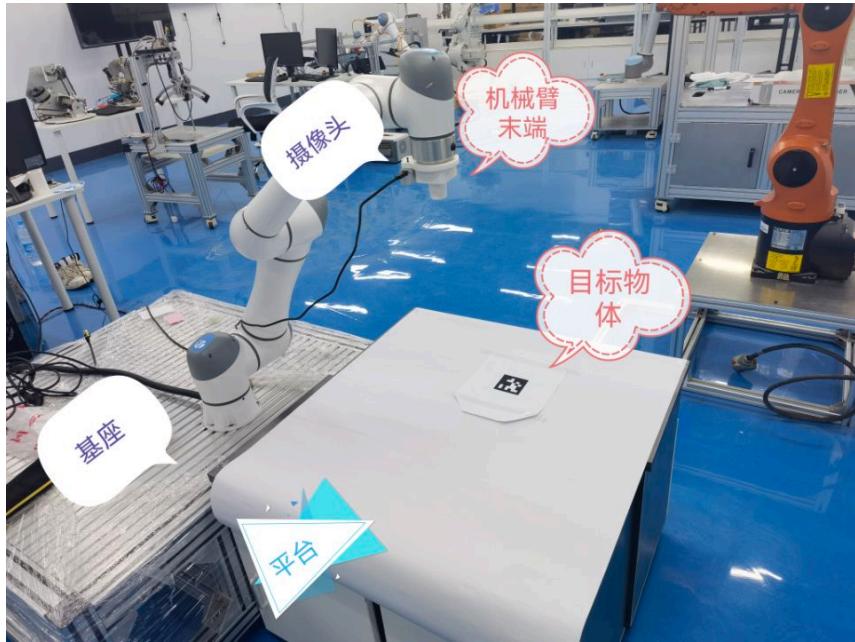


图 2-8 实物环境搭建展示图

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。

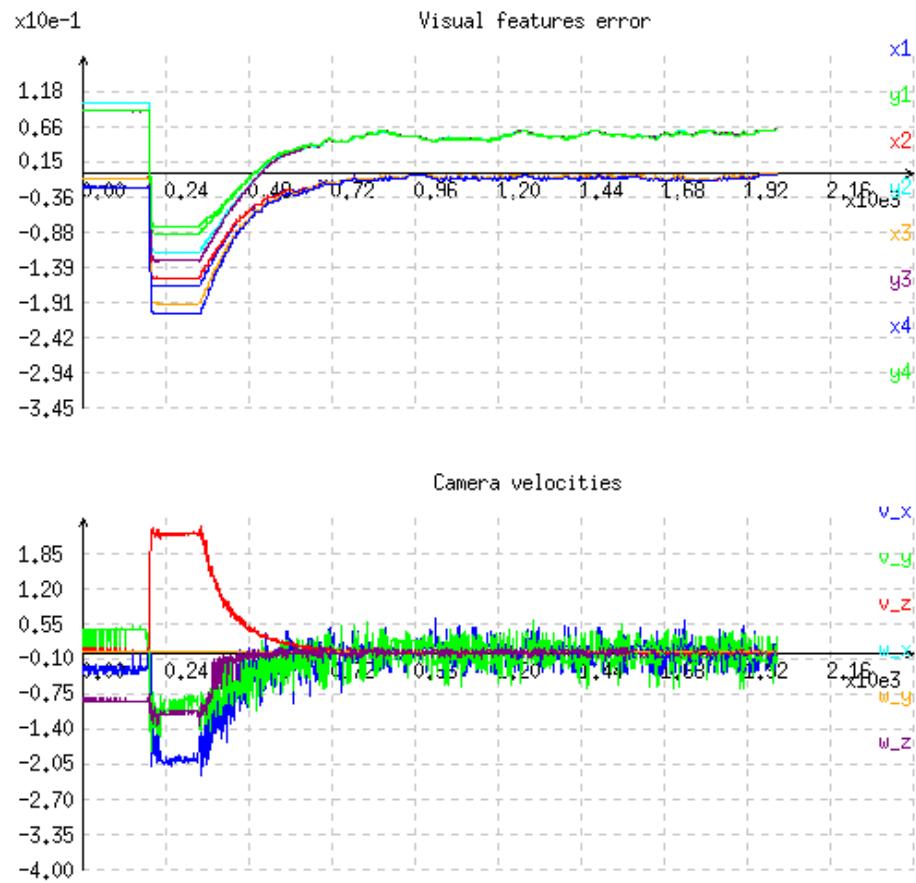


图 2-9 视觉伺服曲线绘制（初始）

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的形式呈现，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3-1所示：



图 3-1 圆柱类点云模型

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^[17]。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3-2和图3-3所示。

图3-2和图3-3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。

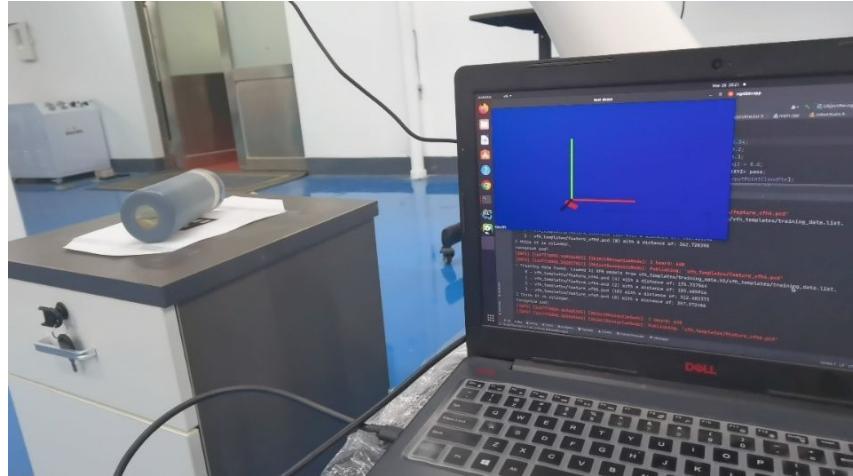


图 3-2 基于模型的点云识别实操图

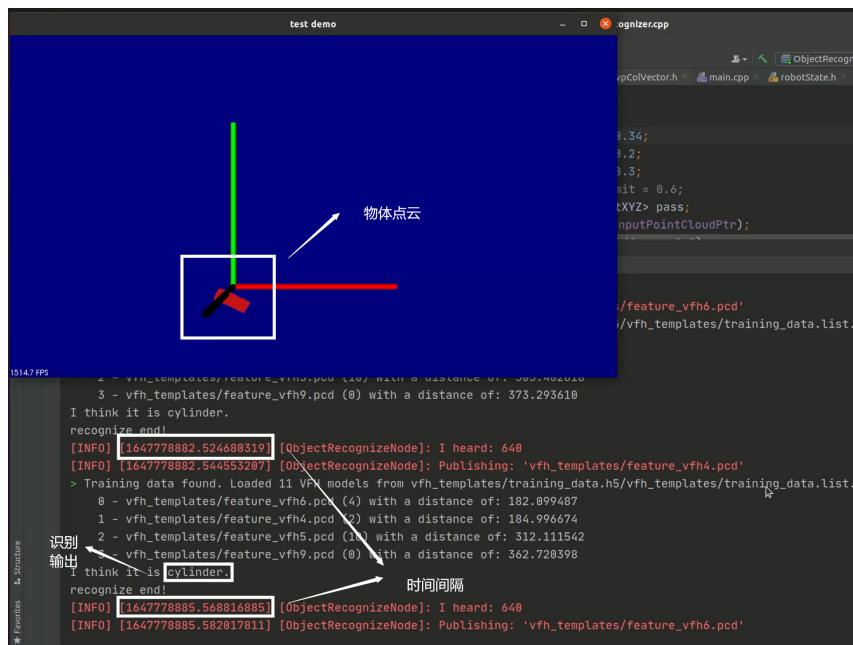


图 3-3 基于模型的点云识别效果图

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一战之力。

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^[18]。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3-4所示。

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

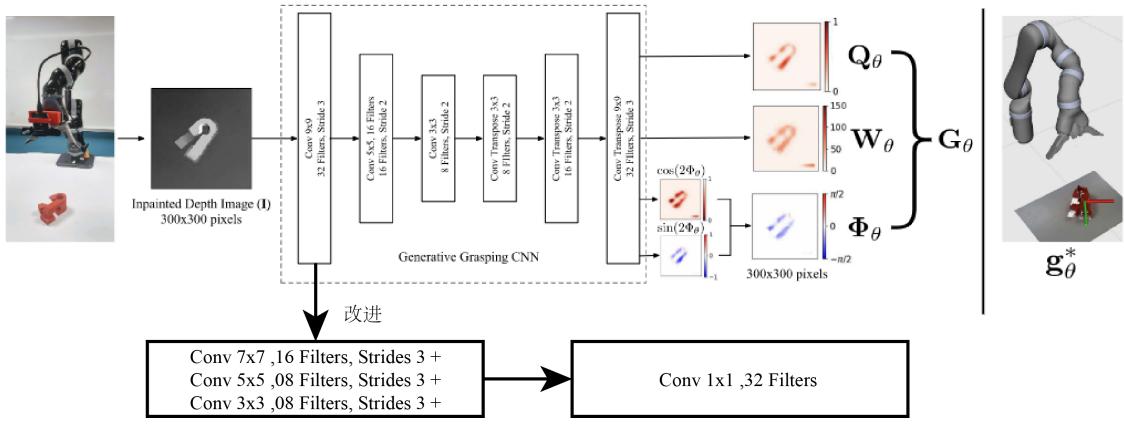


图 3-4 GG-CNN 改进后网络架构

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3-5所示：

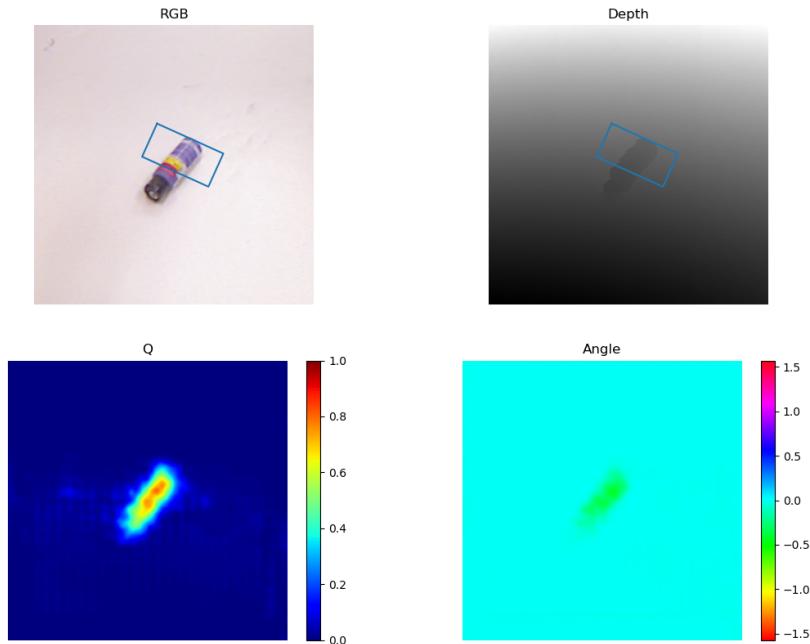


图 3-5 GG-CNN 输出

这里特别解释一下图3-5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

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。选择抓取对象为笔袋，图3-6为实时抓取点生成效果图。



图 3-6 实时抓取点生成

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3-7所示。这会使生成的目标特征不断摆动，导致系统失稳。（由于角度为弧度制过太小，输出时乘了 100 倍）

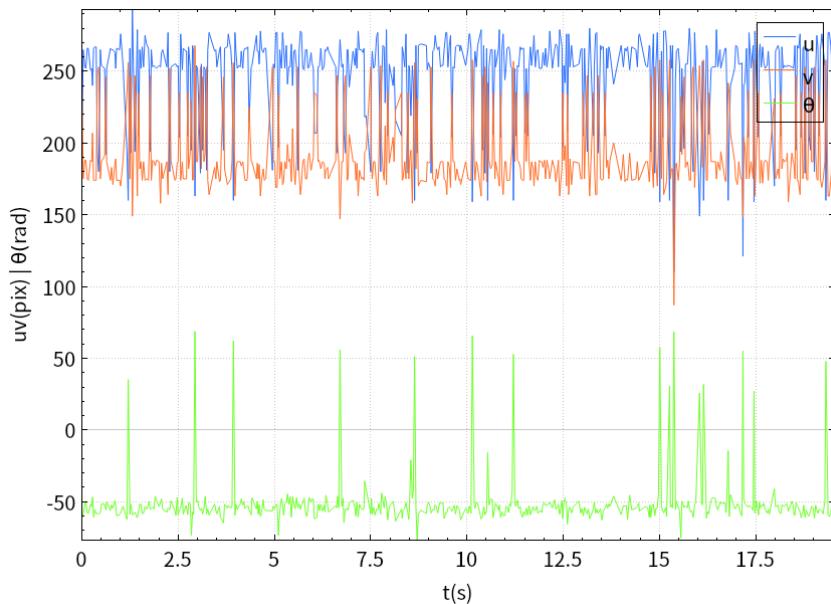


图 3-7 神经网络原始输出

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：

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

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如图3-8所示。得到的抓取中心只有1到2像素点的波动，预测的角度几乎不再波动。

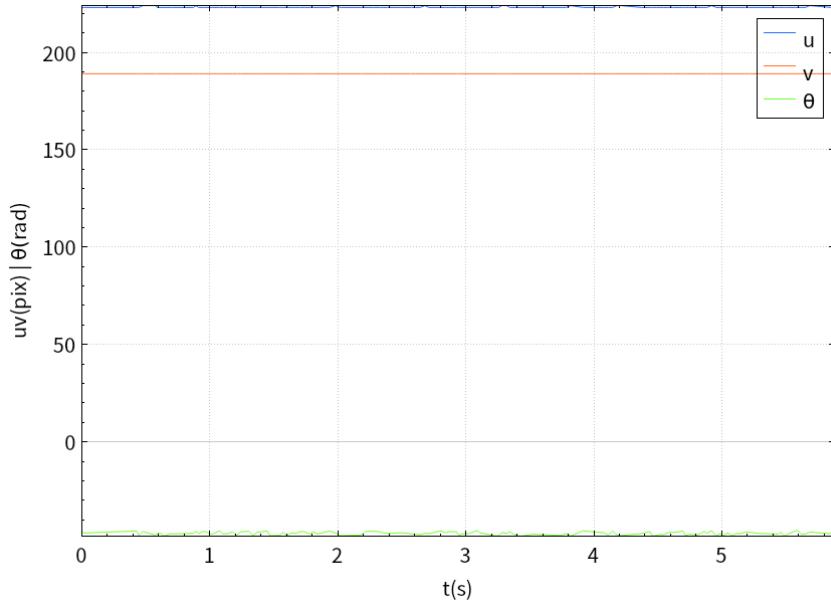


图 3-8 抓取点中心方法替代与角度滤波

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 一阶图像矩来生成抓取点中心。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

第二章、第三章分别解决了控制基础和控制目标自主生成的问题，在对它们研究的设计中无不为本章而服务。本章将正式进行未知物体抓取算法的实现。通常情况下，相机视野里在伺服过程中会出现平台以外的部分，这影响了神经网络的正常工作，将使用图像处理抑制这样的边界环境干扰。然后提出一种基于抓取点和 ORB 特征的 IBVS，在失败中不断累计经验，在该思想上提出一种自定义特征完美实现 GG-CNN 对 IBVS 的内嵌并优化了它的特征点分布。最后对于笔盒和卷纸做了无形状、位姿先验信息的视觉伺服实验，伺服指令平滑而收敛证明了该方法的有效性。

4.2 图像处理抑制边界环境干扰

当抓取目标设置在平台边缘或者抓取角度较大时，视觉伺服过程中，相机视野中会出现平台以外的部分。而神经网络是通过平整的平台上摆放物体的深度图训练的，同时，场景深度的大幅度波动会使检测到的特征深度随之大幅度变化，导致控制指令存在发散的状况。图4-1为出现这种状况时抓取点生成情况，十分明显，生成了错误的抓取点。

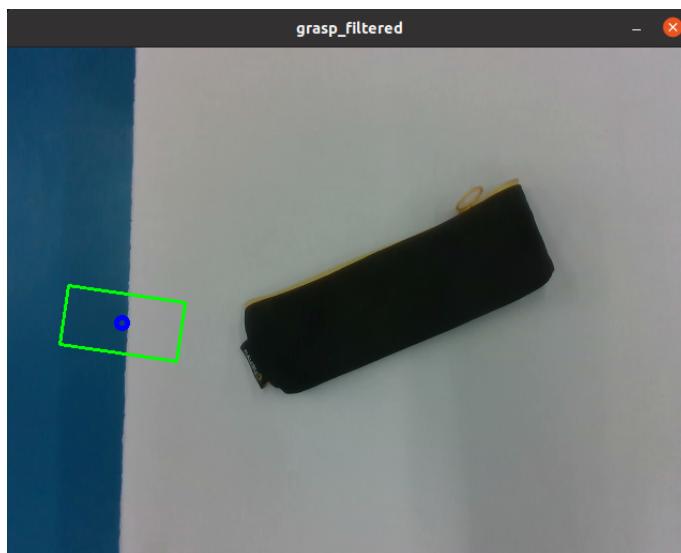


图 4-1 深度突变导致抓取点预测错误

为了解决这个问题，进行一定的图像处理是必要的。目前掌握的信息有

RGB 图和深度图，深度图在之前的实验中已经发现了噪信比大的问题，并且存在许多接受的无效值导致的空洞。首先使用 OpenCV 所带的图像修复函数，选择 Alexandru Telea 发表的“基于快速行进方法的图像修复技术”方法^[19]，对所有无效值区域图像修复，修复前后对比如图4-2所示。左图是未经过修复的深度图像，它左边大块的黑色区域是深度图向彩色图配准时该位置被遮挡而测量得到的无效值，中间空洞处是深度相机测量机制得到的测量无效值。右图是修复后的深度图像，还是会有一些空洞（它是随机出现的，左图的空洞一直会存在），但是它们都出现在深度突变的地方，对目标附近的深度获取是没有影响的。

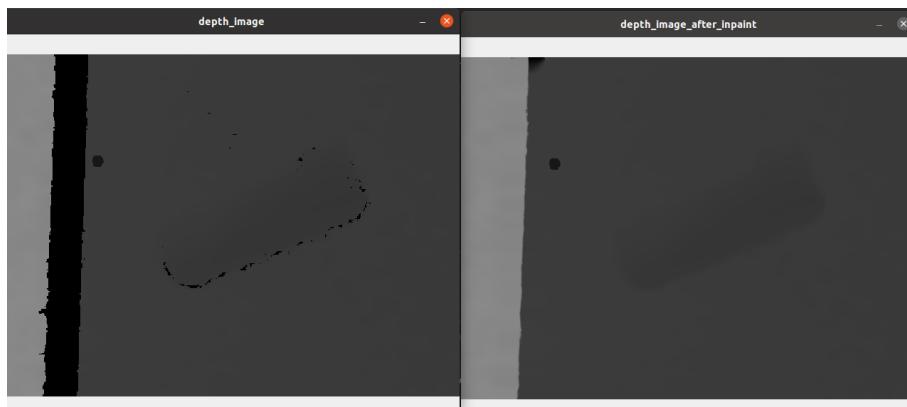


图 4-2 深度图修复前后对比

深度图像经过处理变得十分可靠。注意到平台以外的部分为蓝色的地面对和银色的机械臂基座，接下来的思路就是确定平台位置以及对应的深度，通过判断是否大于平台深度来排除地面和基座。可以通过颜色区分平台和地面，但是机械臂基座和平台颜色相近，二值化后使用颜色中心来确定平台位置并不合适。为此，在对 RGB 图像二值化后，取对应白色区域的深度图像的深度中值作为平台的深度。因为机械臂再怎么运动都不会离开平台过多，离开平台过多情况的出现往往意味着视觉伺服的失败，所以基座对应的 RGB 图像的范围相比平台是更小的，那么取中值能有效地找到平台对应的深度。最后将大于这个深度 0.5cm 位置的深度全部置为平台深度，这样就会把边界的深度都通过平台深度拓展，处理过程见图4-3，图 a 是原始 RGB 图像，图 b 是提取的平台对应像素，图 c 是算法分割出背景后将背景置为白色的 RGB 图像，图 d 是算法分割出背景后将背景置为平台深度的深度图像。

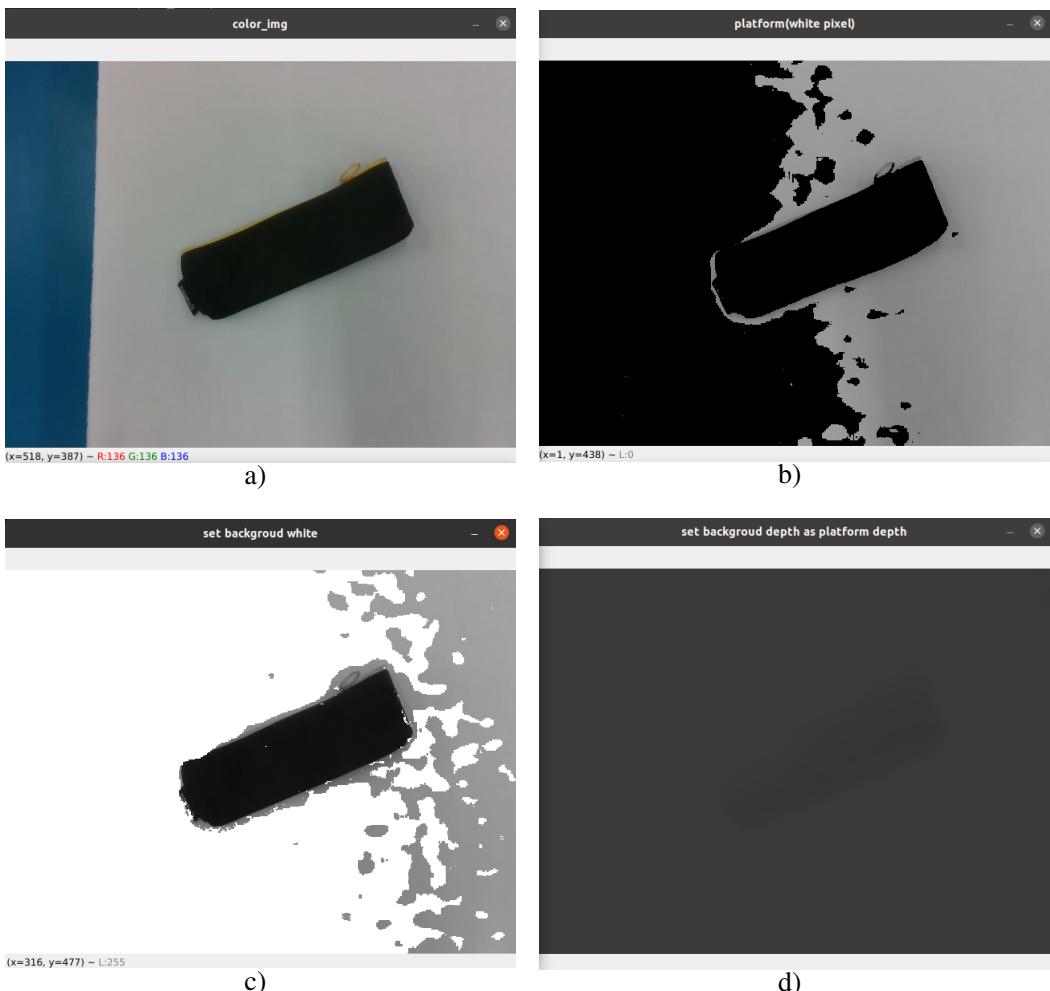


图 4-3 图像处理抑制边缘干扰

4.3 初步思路

4.3.1 ORB 特征的提取与匹配

基于特征点交互矩阵的 IBVS 依赖特征的提取与匹配来完成伺服任务。这需要求特征点的提取与匹配具有很强的稳定性。不仅如此，式2-5不具备特征点顺序的对称性，这要求特征点匹配时需要保证顺序一致，否则速度指令会出现锯齿状的波形。研究中做了一次仿真，选择四个特征点，不断改变特征点对应的顺序，得到的速度指令波形如图4-4所示。指令存在较大波动，所以尽可能保证交互矩阵中的特征点顺序不要改变。

针对以上对特征点性质的要求，选用平移不变性和旋转不变性的 ORB 特征，它基于金字塔结构提取，对于图像的缩放具有一定的鲁棒性。调用 OpenCV 的ORB 特征提取 API，记录每次提取的特征及其描述子并在下次用于匹配。匹配

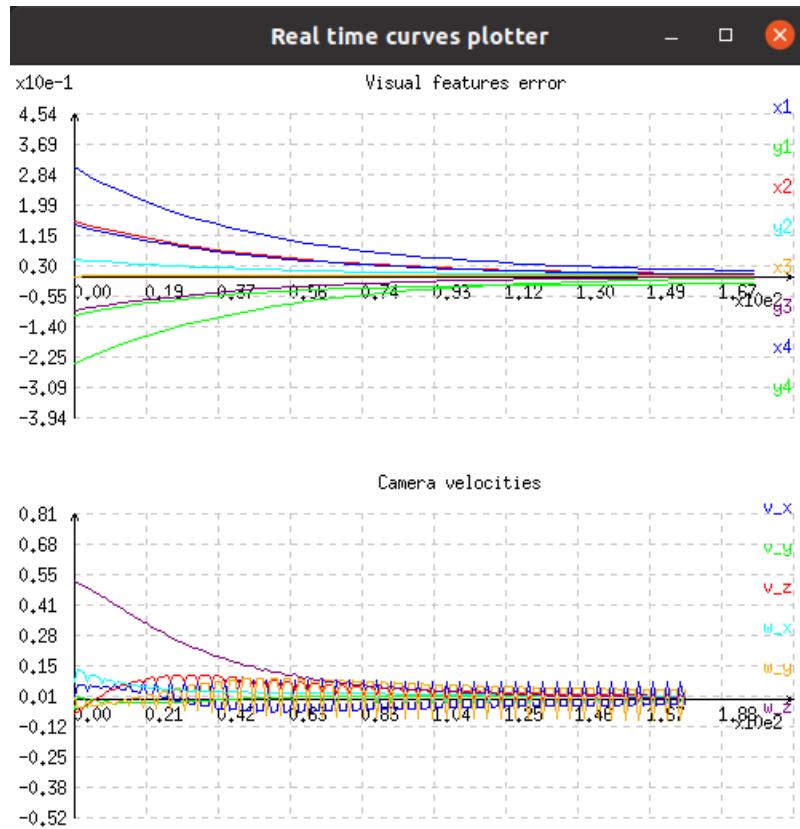


图 4-4 特征变换指令波动

的特征点必须保证描述子汉明距离和两点在图像中的位置相差在一定阈值内，否则认为特征点丢失，从特征点候选中取出来补充，候选特征点个数一直处于 20 个（实际用于运算的特征点数为 4 个），如图4-5中图 a 所示。



图 4-5 ORB 特征点提取与匹配

但是将该算法实时运行时，出现了特征分布收缩的问题，而且仅仅在 30 帧内特征点会彻底聚集在一起。如图4-5中图 b 所示。经过后续的分析，认为这是必然的趋势。纯粹的依靠图像特征提取与匹配，加上距离的约束，那么 4 个几

乎相同的特征点是最稳定的解，它们相似而又相近。

4.3.2 基于抓取点的 IBVS

第三章中实现了一阶图像矩和 GG-CNN 实时输出抓取点。为了能将该信息使用于 IBVS 中，基于这个抓取点和相机一直保持俯视的条件，提出一种基于抓取点的 IBVS。总特征点数还是 4，会时刻计算抓取点附近的特征点并通过抓取点信息自主生成的目标特征，特征点丢失会补充。假设在深度测量器从深度相机向编码器切换后，特征点不会丢失。

首先是目标特征 s^* 自主生成。认为机器人到达了期望位姿时，抓取点应当在图像的中心位置，那么由于相机俯视，此时检测到的特征 s 转化为 s^* 的过程实际上就是一个平面上平移、旋转和放缩的过程。如图4-6所示，这是俯视相机所看到的世界，想把图中相机（三维坐标轴）变换到正对 Apritag 码（模拟抓取点）中心的位置，只需要在 XY 平面上平移，然后按照网络输出的 Z 轴旋转角度沿 Z 轴旋转，最后沿 Z 轴平移到期望深度即可。

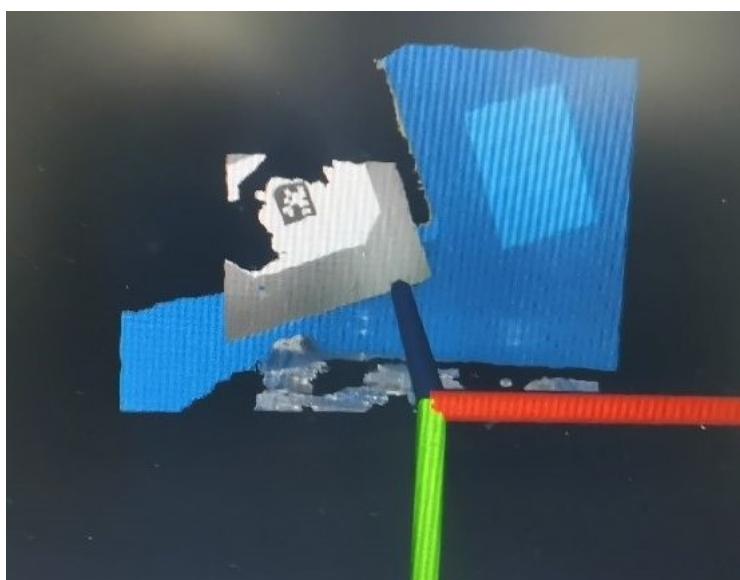


图 4-6 俯视相机成图

由于相机俯视，在图像中的点经历的流程和相机本身是一样的。将抓取点移到图像中心的平移量赋给当前检测到的特征点。设抓取点在图像中的位置为 (u_g, v_g) ，需要旋转的角度为 θ ，设图像大小为 (w, h) ，先将图像坐标系移到按照 $(0, 0)$ 对称的坐标系，平移量为：

$$T_0 = [-w/2, -h/2]^T \quad (4-1)$$

对抓取点平移量：

$$T_1 = (\Delta x, \Delta y) = [w/2 - u_g, h/2 - v_g]^T \quad (4-2)$$

总平移量为：

$$T = T_0 + T_1 = [-u_g, -v_g]^T \quad (4-3)$$

让此时特征点沿原点旋转，旋转矩阵：

$$R = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \quad (4-4)$$

设目标特征 s^* ，目标深度 Z_d 。设当前特征 $s = [u, v]^T$ ，当前特征深度均值 \bar{Z} 。期望特征在图像中的位置可以计算：

$$s^* = \frac{\bar{Z}}{Z_d} (R(s + T) - T_0) \quad (4-5)$$

在笔袋上贴上黑色方框，将方框四个点作为特征来计算目标特征。最终实现效果如图4-7所示，左边蓝色的四个点是根据抓取点生成的四个目标点，在机械臂处于期望位姿时，当前特征点应到处于这个位置。

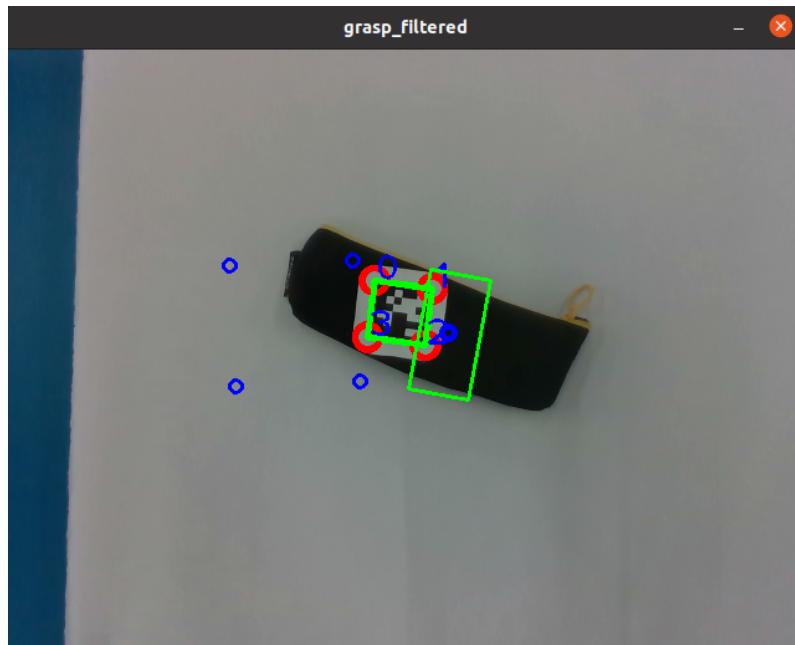


图 4-7 目标特征生成

但非常遗憾的是，该方法跑出的伺服指令曲线是发散的，如图??所示。究其原因在于抓取点生成状态的波动对目标特征的生成影响是很大的。因为机械臂末端处于初始位置时，当前特征比较聚集，生成的目标特征在沿 Z 轴放大后相对误差是很大的。另外，在当前特征的提取和目标特征生成算法下，当前特征和目标特征的位置和旋转角度同时在改变，它们的误差将叠加。

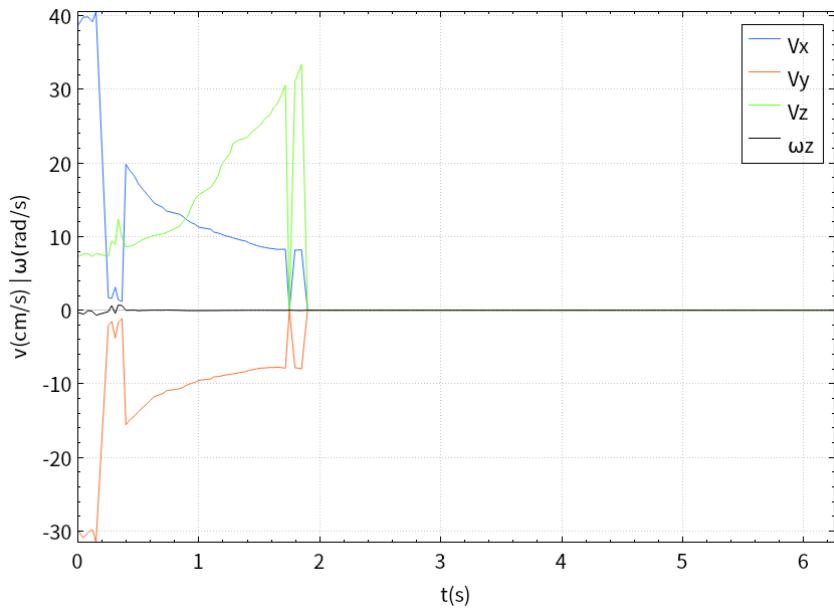


图 4-8 指令发散现象

4.3.3 总结与思考

初步的想法无法实现对于未知物体抓取的任务。但它们的失败为后续给出的解决方法累积了经验。ORB 特征匹配无法持续，是因为特征提取自 RGB 图像的算法中，匹配的最稳定解就是所有特征点相同。而摄像头是移动且具有噪声的，这更是加剧了解向这个方向收敛，所以基于图像的特征提取对于快速伺服是不可行的。通过抓取点计算目标特征来运行 IBVS 的想法也遭遇了困境，它失败的核心原因在于当前特征点和目标特征点的位置和角度同时在改变，误差会累加。另外，没有把控特征点的分布，特征点具有的信息不全面，所以伺服效果不会好。

4.4 一种基于抓取点的自定义特征

4.4.1 方法概述

图像提取特征不可取，既然知道抓取点的位置，通过人为定义特征点相对于抓取点的位置分布可以保证特征提供位置信息的全面性。固定目标特征的位置，保证只有特征的位置在改变，这样不会出现严重的误差叠加现象。选取的当前特征和目标特征分布如图4-9所示。当前特征相对于目标特征的改变在于它经历了旋转和放缩。旋转矩阵 R 由 GG-CNN 提供，并且表达式如式4-4，放缩因子 m 为期望的与物体表面的深度 h_d 和当前与物体表面深度 \bar{h} 的比值。

可以通过深度相机测量当前特征点对应的深度，那么现在只剩下目标特征

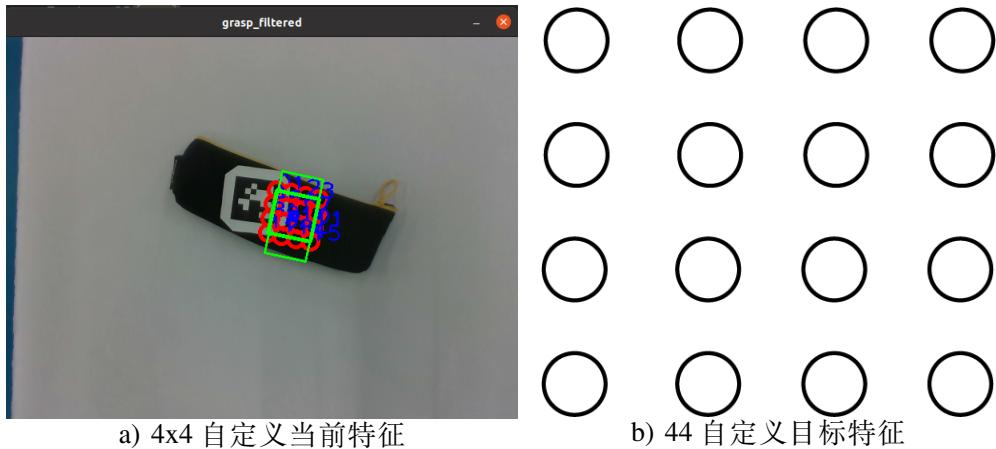


图 4-9 4x4 自定义当前特征

深度取何值的问题。将使用一种迭代的方式求取物体表面深度 \bar{h} ，结合当前特征深度 h 与这个值的差值以及预设的机械臂在期望位姿时距离物体表面的深度 h_d 求取目标特征深度。目标特征深度 h^* 表达式如下：

$$h^* = h - \bar{h} + h_d \quad (4-6)$$

可以求得当前特征位置：

$$s = mR(s^* + T_0) - T \quad (4-7)$$

这样特征和目标特征的信息就完整了，将它们给予 IBVS 控制器即可完成伺服。称这种特征叫做基于抓取点的自定义特征，因为这个特征分布排列是自定义的。

4.4.2 目标表面计算

为了求取目标特征深度，必须知道物体表面的深度，距离物体表面一定距离时给机械臂下达停止指令，准备抓取。但是对于一个未知形状的物体，它可以是圆锥这种不存在平面表面的物体，单取物体对应像素处的深度的平均值来作为物体表面深度是不合适的。所以使用一种迭代的方式，不断滤除大于当前物体深度平均值的像素点，取剩余像素点深度的平均值继续滤除。迭代示意如图4-10所示。算法效果相当于不断通过平均深度平面截断目标物体，获得平面以上的部分继续求平均深度，然后重复以上步骤。

实验验证重复两次是合适的，继续迭代所获得的物体表面深度改变在 1mm 左右，已经到达深度相机的深度分辨率了。另外，因为滤除次数过多会导致计算量增大，剩余点数较少，受噪声的影响变大。对于卷纸的迭代前后物体像素对比如图4-11所示。该图将笔袋竖着放置，左图为迭代前经过图像处理得到的

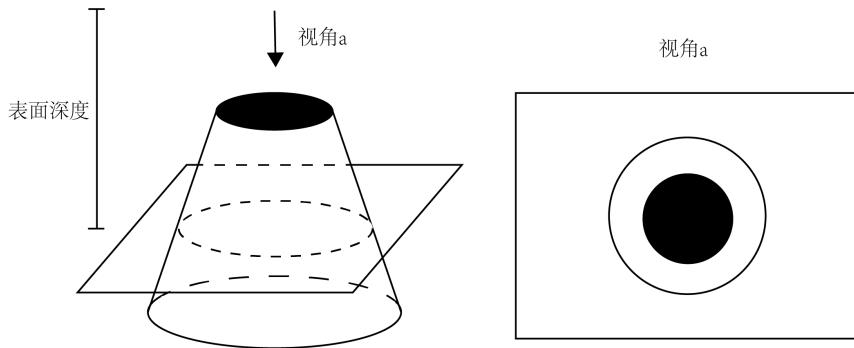


图 4-10 迭代示意图

笔袋对应的像素，右图为迭代后对应的像素。经过迭代后，用于求平均深度对应的像素只有笔袋的上半部分。获得的物体表面深度为 0.4464m，由卷尺测量为 0.4460m，毫米级的误差证明了算法的有效性。



图 4-11 迭代求前景前后对比

在运行 IBVS 程序，时遇到了 Z 轴速度指令一直接近 0 的问题，不论抓取对象是笔袋还是卷纸。指令曲线如图??所示。经过反复实验发现只要特征分布恰好全在现实世界的一个平面上且旋转角度偏差很小的时候就会导致这个问题。对于笔袋，当前设定的特征分布相对聚集，所有特征点都在笔袋上；对于卷纸，当前设定的特征分布相对松散，所有特征点都在平台上，所以需要优化当前特征分布。

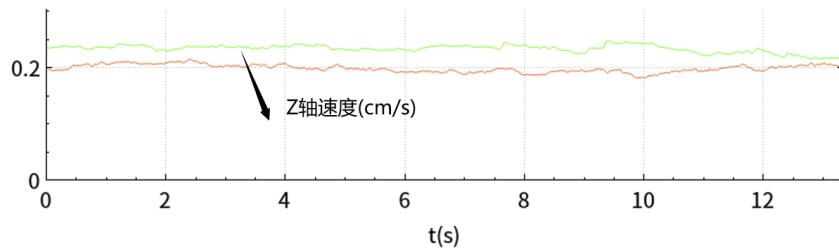


图 4-12 Z 轴速度接近 0 的情况

4.4.3 优化特征分布与 IBVS 试验

当前特征需要涵盖平台、物体和物体细节的位置信息，最终设计的特征分布如图所示。这样的设计保证了特征点能针对任意常见物体获取更多位置信息，Z 轴速度指令回归正常。将笔袋作为抓取对象，执行 IBVS 程序，伺服效果如图4-14所示。

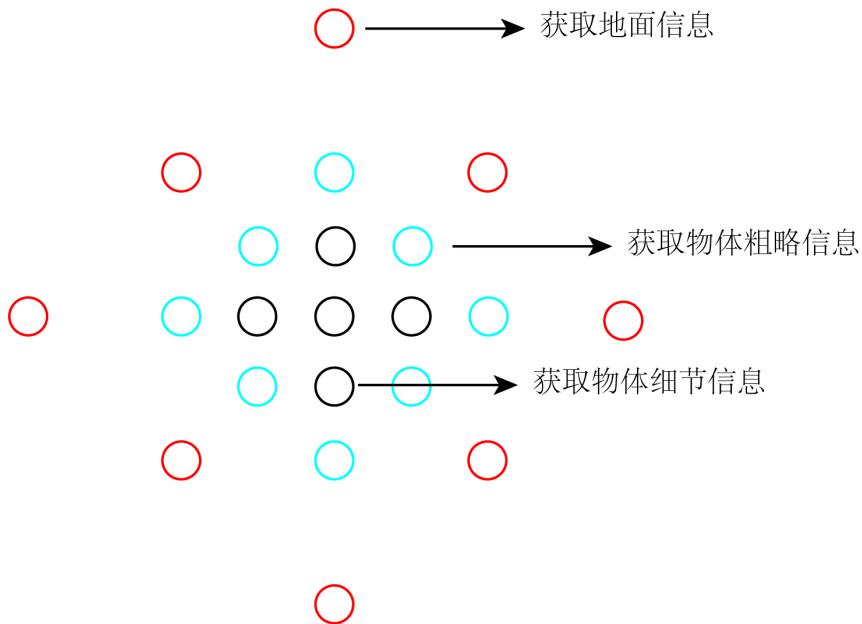


图 4-13 优化自定义目标特征

注意到在图4-14 b)中，整个伺服过程所消耗的时间长达 25s，此时的控制系统只能保证最基础的稳定性而已，在快速性方面有待提高。另外值得一提的是，所提出的特征分布不一定拘泥于 GG-CNN 算出的抓取点，对于俯视二指爪抓取物体的问题，任意算法给出一个抓取点，它都可以当一个插件挂上去，帮助运行 IBVS。

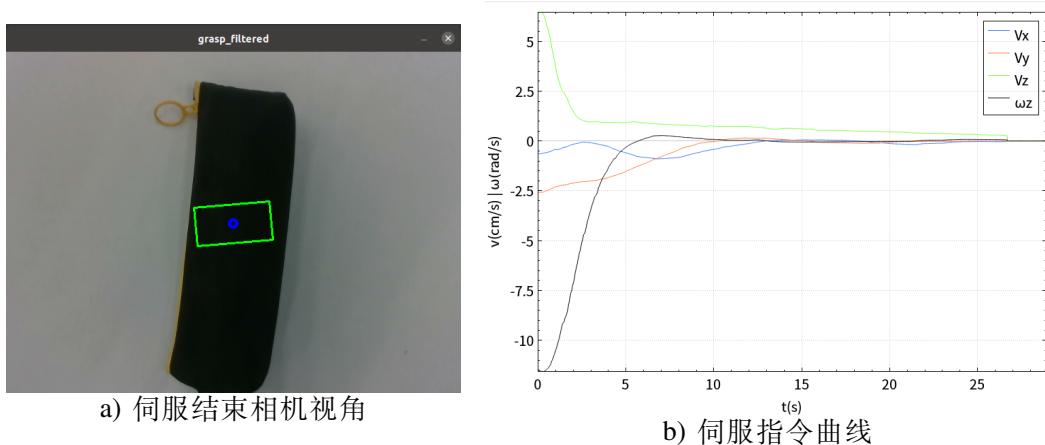


图 4-14 优化特征分布后伺服效果

4.5 本章小结

本章首先通过图像处理解决了边界环境干扰问题，然后结合了前两章的研究成果提出了一种基于抓取点的IBVS，但是在实现过程中遭遇了失败。结合失败的经验，在当前研究成果上改进特征提取方法，提出一种基于抓取点的自定义特征，避免了特征和目标特征都在改变导致的误差叠加问题的同时，特征的排列顺序和匹配是可靠性问题也得以解决。

第 5 章 IBVS 控制律优化

5.1 引言

之前的研究已经完成了机械臂基于视觉伺服的未知物体抓取算法，但是伺服性能是较差的。优秀的控制律会让视觉伺服系统精确、快速地收敛，这不仅会让抓取算法具有实用性，而且还能配合抓取算法提升系统抓取物体的成功率。本章将建立 IBVS 控制模型，使用 PD 控制器和惯性环节提升伺服性能，使用滑摸控制器让系统对深度测量的不确定性具有鲁棒性，然后基于 Lyapunov 直接法证明系统稳定性。最后注意到特征偏差关联速度不一致性，利用特征偏差之间的相对大小改写交互矩阵，实现了一种自适应控制算法，使所有特征与最快收敛的特征同时收敛，提升了收敛速度。

5.2 控制模型建立

5.3 PD 控制器与惯性环节

5.4 滑模控制器

5.5 系统稳定性

5.6 特征偏差关联速度不一致性

5.7 本章小结

本章在当前抓取算法下对视觉伺服系统的控制律进行优化。PD 控制器能有效提升伺服性能，滑模控制器一定程度抑制了深度相机对深度测量的不确定性，然后基于 Lyapunov 直接法证明系统稳定性。针对出现的特征偏差关联速度不一致性，提出了一种自适应控制算法，通过使所有特征与最快收敛的特征同时收敛，带来了收敛速度的提升。

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 OY 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，目标摆放位置和对应姿态如表所示

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^[20] 进行比较。

6.3 实验验证

6.3.1 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：

表 6-1 符合研究生院绘图规范的表格

D (in)	P_u (lbs)	u_u (in)	β	G_f (psi.in)
5	269.8	0.000674	1.79	0.04089
10	421.0	0.001035	3.59	0.04089
20	640.2	0.001565	7.18	0.04089
D (in)	P_u (lbs)	u_u (in)	β	G_f (psi.in)

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：描述

一下

表 6-2 符合研究生院绘图规范的表格
Table 6-2 Table in agreement of the standard from graduate school

D (in)	P_u (lbs)	u_u (in)	β	G_f (psi.in)
5	269.8	0.000674	1.79	0.04089
10	421.0	0.001035	3.59	0.04089
20	640.2	0.001565	7.18	0.04089

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 6-3 符合研究生院绘图规范的表格
Table 6-3 Table in agreement of the standard from graduate school

D (in)	P_u (lbs)	u_u (in)	β	G_f (psi.in)
5	269.8	0.000674	1.79	0.04089
10	421.0	0.001035	3.59	0.04089
20	640.2	0.001565	7.18	0.04089

表格反映了。。。

6.3.2 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。。。

6.4 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

科技日新月异，我国综合国力在飞速提升，机器人相关技术也在迅猛发展。工业、服务智能化对机器人自适应能力提出了苛刻的要求，视觉伺服已经几乎是智能机器人不可缺少的部分。本文以手在眼机械臂视觉伺服系统为研究对象，以抓取未知形状随机位姿的物体为目标展开研究，旨在面对多样的抓取对象，与其它类似方法比较也能拥有更高的抓取成功率。在研究中提出一种基于抓取点的自定义特征，保证特征匹配的可靠性和特征排序不变性。在当前算法基础上给出适合它的 IBVS 控制律，针对特征偏差对末端速度指令关联的不对称性问题和噪声问题提出一种自适应控制算法。最后通过实物实验验证提出的基于抓取点生成网络和 IBVS 的未知物体抓取算法的高抓取率、高伺服性能。

本文主要研究内容如下：

- (1) 针对未知物体抓取任务搭建 IBVS 系统模型，其中包括系统坐标系和视觉坐标系的建立。针对以上建立的模型和基于特征点交互矩阵的 IBVS 原理，实现 IBVS 系统仿真，并实现数据曲线记录和相机轨迹绘制功能方便调试。最后搭建实物平台，在实物上成功运行 IBVS 算法。
- (2) 研究了基于模型的点云识别与配准和抓取点生成网络 (GG-CNN) 两种生成伺服目标的方法，通过比较它们的优劣选择了后者作为伺服目标生成方案。最后采取滤波等方式抑制了神经网络的输出波动。
- (3) 通过图像处理解决各种环境干扰问题。将抓取点生成网络运用到 IBVS 中，提出一种基于抓取点的自定义特征并根据伺服效果优化了特征分布，实现了未知物体抓取算法并在实物上成功运行。
- (4) 针对当前伺服性能差的问题，建立控制模型，根据伺服曲线给出 IBVS 控制律的改善方法。相比于原始的伺服指令曲线和特征偏差曲线，改进后的曲线更加理想且光滑。对于出现的特征偏差对末端速度指令关联的不一致性问题和噪声问题提出一种自适应控制算法。
- (5) 设计了抓取实验，通过多样的物体、多样的位姿的抓取实验验证提出的未知物体抓取算法的有效性。与相关工作进行比较，证明当前算法的高自适应能力和高伺服性能。

综上所述，本文对基于视觉伺服的机械臂抓取未知形体、任意位姿物体的研究取得了一定的进展，但根据当前伺服的实际情况可知在未来的研究中仍然

有许多问题需要去解决。那些需要继续研究的方面将在下文列出。

(1) 本文研究中使用了 GG-CNN 作为抓取点生成的方式之一，但在实时抓取点生成中发现抓取点位置波动严重到无法使用，所以使用了图像一阶矩代替成为获取抓取点中心位置的方案，仅使用了 GG-CNN 的抓取点偏转角信息。训练网络时填充自己的数据集也许能部分解决问题，但不能从根本上解决该网络对深度的平移、旋转不变性的问题。一个更好的方式配合基于抓取点的自定义特征实现 IBVS，能让机械臂的自适应能力变得更好。

(2) 本文提出的于抓取点的自定义特征仅能用于相机一直俯视向下的伺服系统，这对该方法的使用带来了一定的局限性。因此，如果能借助机械臂编码器信息，通过优化方法使该方法在不借助外参的情况下在六自由度速度指令的伺服系统中使用，该方法就能作为插件广泛应用于各个 IBVS 系统中了。

(3) 尽管本文在对于视觉伺服控制律优化的研究结果是好的，但还是不够深入的。仅从时域的角度对控制律进行优化，可能会遗漏关于优化方法上的重要信息。如果能从频域角度分析指令和对象然后设计控制器可能会为这个伺服系统带来更高的伺服性能。

参考文献

- [1] 徐鑫莉, 韩啸, 刑玉龙, et al. 工业机器人视觉伺服控制系统设计[J]. 伺服控制, 2015(8): 79-81.
- [2] 陶波, 龚泽宇, 丁汉. 机器人无标定视觉伺服控制研究进展[J]. 力学学报, 2016, 48(4): 767-783.
- [3] Tsuchida S, Lu H, Kamiya T, et al. Characteristics based visual servo for 6DOF robot arm control[J]. Cognitive Robotics, 2021, 1: 76-82.
- [4] 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.
- [5] 贾丙西, 刘山, 张凯祥, 等. 机器人视觉伺服研究进展: 视觉系统与控制策略[J]. 自动化学报, 2015, 41(5): 861-873.
- [6] Breedveld P, Stassen H, Meijer D, et al. Theoretical background and conceptual solution for depth perception and eye-hand coordination problems in laparoscopic surgery[J]. Minimally invasive therapy & allied technologies, 1999, 8(4): 227-234.
- [7] Eisenberg D, Vidovszky T J, Lau J, et al. Comparison of robotic and laparoendoscopic single-site surgery systems in a suturing and knot tying task[J]. Surgical endoscopy, 2013, 27(9): 3182-3186.
- [8] 卢钰. 基于双目磁锚定手术机器人无标定视觉伺服控制研究[D]. 哈尔滨: 哈尔滨工业大学, 2016.
- [9] Chaumette F, Hutchinson S. Visual servo control. I. Basic approaches[J]. IEEE Robotics & Automation Magazine, 2006, 13(4): 82-90.
- [10] Chaumette F, Hutchinson S. Visual servo control. II. Advanced approaches [Tutorial][J]. IEEE Robotics & Automation Magazine, 2007, 14(1): 109-118.
- [11] Chaumette F. Image moments: a general and useful set of features for visual servoing[J]. IEEE Transactions on Robotics, 2004, 20(4): 713-723.
- [12] Liu C, Chen R, Xu J, et al. Set space visual servoing of a 6-dof manipulator[C] // 2017 IEEE International Conference on Robotics and Automation (ICRA). 2017: 4422-4428.
- [13] Tokuda F, Arai S, Kosuge K. Convolutional neural network-based visual servoing for eye-to-hand manipulator[J]. IEEE Access, 2021, 9: 91820-91835.

- [14] Varley J, Weisz J, Weiss J, et al. Generating multi-fingered robotic grasps via deep learning[C] // 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS). 2015: 4415-4420.
- [15] Adrian N, Do V-T, Pham Q-C. DFBVS: Deep Feature-Based Visual Servo[J]. arXiv preprint arXiv:2201.08046, 2022.
- [16] Rauch C, Ivan V, Hospedales T, et al. Learning-driven coarse-to-fine articulated robot tracking[C] // 2019 International Conference on Robotics and Automation (ICRA). 2019: 6604-6610.
- [17] 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.
- [18] 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.
- [19] Telea A. An image inpainting technique based on the fast marching method[J]. Journal of graphics tools, 2004, 9(1): 23-34.
- [20] 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识别算法研究》，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: 王泽飞 日期: 2022 年 5 月 23 日

致 谢

衷心感谢导师高会军教授和林伟阳副教授对本人的精心指导。他们的言传身教将使我终生受益。感谢姜易木、刘晨璐师兄对我搭建视觉伺服实验环境的帮助和刘海超师兄给我关于特征设计上很棒的启发。感谢霍鑫老师对我控制设计方面的指导和室友武博涵对人机交互界面设计上的大力支持。和一些朋友之间的点点滴滴的交谈中有时也能收获很不错的想法，感谢你们！最后感谢父母和女朋友曹丹琳在我完成毕业设计过程中精神上的莫大支持！

感谢大家，你们每个人的帮助也许只是只言片语，也许重如泰山，但正是这些组成了现在我十分满意的“作品”。另外，我要毕业啦!!!