

# 面向机器人控制的直接视觉伺服技术发展综述

黄梦洁, 叶 磊<sup>†</sup>, 易凡骁, 王 千

(国防科技大学 智能科学学院, 湖南 长沙 410073)

**摘要:** 视觉伺服作为一种利用图像信息来实现机器人闭环控制的方法, 在移动机器人导航控制和机械臂高精度控制等方面有广泛的应用。相较于基于局部特征的经典视觉伺服方法, 直接视觉伺服技术利用了更多的图像全局信息, 在控制系统的鲁棒性和收敛性上有较大优势, 因此本文重点从视觉处理和控制器设计两部分介绍直接视觉伺服技术。本文首先介绍了经典视觉伺服框架, 然后研究了直接视觉伺服技术中的视觉特征描述方法、特征相似性度量方法, 并对比了不同特征描述方法在收敛域和运算量等性能方面的差异, 之后研究了直接视觉伺服中基于非线性优化等优化策略的控制器设计方法, 最后展望了深度学习在直接视觉伺服中的作用。

**关键词:** 机器人控制; 视觉伺服; 直接视觉伺服; 图像特征描述方法; 相似性度量; 优化控制策略

**引用格式:** 黄梦洁, 叶磊, 易凡骁, 等. 面向机器人控制的直接视觉伺服技术发展综述. 控制理论与应用, 2025, 42(9): 1681 – 1699

DOI: 10.7641/CTA.2024.30248

## A review of the development of direct vision servo technology for robot control

HUANG Meng-jie, YE Lei<sup>†</sup>, YI Fan-xiao, WANG Qian

(College of Intelligence Science and Technology, National University of Defense Technology, Changsha Hunan 410073, China)

**Abstract:** Visual servo is a method that uses image information to realize closed-loop control of robots, which has a wide range of applications in mobile robot navigation control and high-precision control of robotic arms. Compared with the classical visual servo method based on local features, direct visual servo technology makes more use of image global information, which has great advantages in the robustness and convergence of the control system. So this paper focuses on direct visual servo technology from two parts: visual processing and controller designing. This paper first introduces the classical visual servo framework, then studies the visual feature description method and the feature similarity measurement method in direct visual servo technology, and compares the differences in the performance of different feature description methods in convergence domain and computing amount, and then studies the controller design methods based on optimization strategy like nonlinear optimization in direct visual servo, and finally looks forward to the role of deep learning in direct visual servo.

**Key words:** robot control; visual servo; direct visual servo; description methods of image features; similarity measures; optimize control strategies

**Citation:** HUANG Mengjie, YE Lei, YI Fanxiao, et al. A review of the development of direct vision servo technology for robot control. *Control Theory & Applications*, 2025, 42(9): 1681 – 1699

## 1 引言

随着机器人技术的迅猛发展, 具备感知能力的智能机器人已是当今的主流趋势。在诸多传感器中, 视觉传感器因其信息量大、适用范围广、非接触性等特点得到了广泛应用。自 Shirai 和 Inoue<sup>[1]</sup>首次利用视觉反馈回路来提高机器人控制的准确性以来, 视觉控制方面涌现了大量的研究工作<sup>[2–7]</sup>。将视觉传感器引入机器人控制系统, 不仅可以增加机器人对环境的自适

应性和灵活性, 还可以提高了机器人控制的精确性。由此, 催生了视觉伺服技术的出现。

视觉伺服控制涉及计算机视觉、机器人技术、控制理论等多个领域, 国内外学者的研究一直在深入。国外 Hutchinson 等<sup>[2–4]</sup>的经典论文引导了视觉伺服控制的研究, 国内方勇纯<sup>[5]</sup>、贾丙西<sup>[6]</sup>、徐德<sup>[7]</sup>、张雪波<sup>[8]</sup>等人也针对视觉伺服控制从不同的角度进行了详尽的论述。

收稿日期: 2023–04–20; 录用日期: 2024–01–02.

<sup>†</sup>通信作者。E-mail: yelei@nudt.edu.cn; Tel.: +86 13308490108.

本文责任编辑: 辛景民。

在以往的视觉伺服技术中,视觉系统和机器人控制系统大多以开环方式结合,即根据图像间接计算机器人的当前位姿或相对目标物体的相对位姿,利用位姿信息驱动机器人完成伺服任务。该方法实际控制的精度直接取决于视觉传感器和机器人末端执行器的精度,但在这个过程中存在如下几个问题:

首先,该过程需要进行大量标定工作,易出现标定过程繁琐、标定成本高、标定结果影响控制精度等问题,因此出现了无需预先标定的控制方法,即直接基于图像来定义系统误差。其次,以往的视觉伺服方法大多需要对图像进行局部特征抽取,而特征提取的质量会直接影响控制精度,因此出现了不进行局部的特征提取的视觉伺服方法,使用全局图像特征进行直接视觉伺服控制<sup>[9-12]</sup>。在高精度雕刻、芯片加工等典型伺服场景也奏效的场景下,直接视觉伺服能够满足更高的精度需求,此外,针对服务航天器<sup>[13]</sup>也开始了直接视觉伺服的相关研究。

本文基于以往视觉伺服的发展概况,考虑到使用局部特征的视觉伺服来控制机器人存在大量问题,鉴于使用全局特征的直接视觉伺服技术能为部分问题

的解决提供思路,特在此对面向机器人控制的直接视觉伺服技术进行如下综述。以机器人控制为背景,以求解问题的流程为文章脉络进行展开。主要包括全局图像特征描述方式及相应的交互矩阵的介绍、图像间的特征相似性度量手段与成本函数构建的介绍、以求解优化问题为目的及最终控制策略方法的介绍。

在介绍直接视觉伺服技术之前,首先在第1节简要介绍了3种经典的视觉伺服框架,并根据是否使用图像的全局特征,引出了对直接视觉伺服技术的讨论。第2节和第3节分别从视觉以及控制两方面对直接视觉伺服的相关优缺点展开介绍。第2节主要介绍了直接视觉伺服使用的图像特征描述方法。首先对几种基本的描述方式进行简要分析,随后针对发展较迅速的基于光度变换的特征描述方法进行了分类整理,并在视觉部分的最后介绍了多种度量图像间的特征相似性的方法。第3节主要介绍了直接视觉伺服使用的控制策略。包括几类非线性优化方法的介绍以及基于粒子滤波的控制策略。最后,对直接视觉伺服技术的发展进行了总结与展望。图1列出了本文针对直接视觉伺服技术进行综述的主体框架。

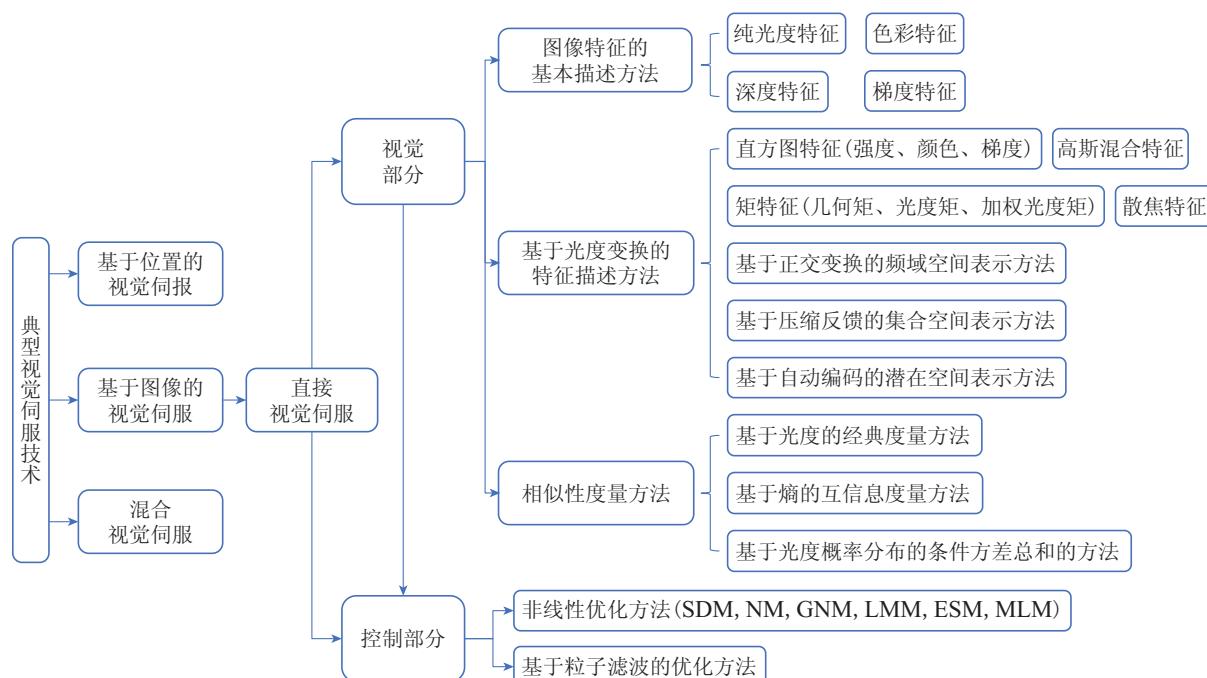


图1 直接视觉伺服技术的发展框架

Fig. 1 A framework for the development of direct vision servo technology

## 2 典型视觉伺服与直接视觉伺服

视觉伺服系统主要包括视觉处理模块和控制器模块,视觉处理模块用于检测机器人当前状态与期望状态的误差,控制器模块根据测量到的误差产生控制信号使误差逐步收敛<sup>[2-4]</sup>。和典型的视觉伺服系统相比,直接视觉伺服技术<sup>[9-12]</sup>同样基于上述框架,其差别在于使用的特征描述方式及求解控制问题的方式。

### 2.1 典型的视觉伺服框架

视觉伺服(visual servo, VS)作为一种在机器人控制领域广泛使用的闭环控制方法,其反馈信号来源于图像。根据反馈信号的类型可将其分为:基于位置的视觉伺服(position-based vs, PBVS)、基于图像的视觉伺服(image-based vs, IBVS)以及综合上述两者优点提出的混合视觉伺服(hybrid vs, HVS)<sup>[5-7]</sup>。

表1<sup>[14-21]</sup>展示了3类视觉伺服框架的优缺点对比 及后续关于直接视觉伺服(direct vs, DVS)的对比.

表 1 3种视觉伺服框架的比较

Table 1 Comparison of three visual servo frameworks

经典视觉伺服框架与DVS	优点	缺点
PBVS <sup>[14-16]</sup>	直接在3D空间中操作更直观 分开研究视觉重建和机器人控制增加了可研究性 机器人的运动轨迹更合理 保证系统的全局渐近稳定	标定任务多增加过程的复杂度 标定影响控制精度 精度受机器人模型影响大 计算要求高计算量大 难以实时估计位姿
IBVS <sup>[17-18]</sup>	具有较强的鲁棒性 通常能保证物体满足FOV约束 静态定位精度较高 受标定影响小 标定任务少只需要校准视觉和机器人	不能保证整个任务空间收敛 存在雅可比矩阵奇异性和局部极小 对于大幅度旋转伺服效果差
HVS <sup>[19-21]</sup>	提高系统的稳定性和收敛性 能保证一定的控制精度	不能保证物体满足FOV约束 分解单应性时解不唯一
DVS <sup>[17-18]</sup>	具有较强的鲁棒性 通常能保证物体满足FOV约束 受标定影响小, 定位精度较高 稳定性强	收敛域较小 存在雅可比矩阵奇异性和局部极小 实时性需要提升

### 2.1.1 基于位置的视觉伺服

PBVS的基本框架如图2所示, 首先需要利用图像信息、相机模型以及目标结构模型来进行三维重构, 再利用多视图几何、三焦张量等方法来估计机器人的位姿信息, 进而结合设定的位姿信息构造误差信号形成闭环反馈控制. 由于位姿误差信号定义在3D笛卡尔空间下, 因此也称3D视觉伺服.

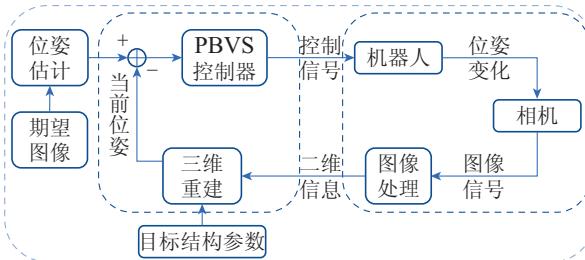


图 2 基于位置的视觉伺服

Fig. 2 Position-based visual servo (PBVS)

在PBVS进行位姿估计的过程中, 需要对视觉与机器人系统进行标定, 其标定内容包括但不限于相机标定、机器人标定、机器人与相机相对位置标定等, 标定工作的存在不仅增加了视觉伺服过程的复杂度, 而且相机的标定误差和机器人的模型误差对控制精度影响非常大, 而基于PBVS的控制器又都基于相对位姿进行设计的, 因此对于力求实现高精度控制的机器人而言往往难度较大. 同时随着机器人对控制精度要求的提高, 对位姿准确度的计算要求也会提高, 相应的计算量也会随之增加, 这反过来又成为了机器人实时

姿态估计的一个挑战.

虽然PBVS中以开环形式结合视觉与控制的方式, 导致控制精度受影响较大, 但这也为视觉重建的研究提供了更多可能性. 由于PBVS能实现3D空间下的直接控制, 因此伺服过程更直观的同时也能更合理地控制机器人的运动轨迹, 最重要的是该方法能够保证系统的全局渐近稳定. 因此在机器人控制领域, PBVS得到了广泛应用. 在移动机器人上, 曾用于移动机器人的路径规划<sup>[14]</sup>、自适应视觉伺服镇定控制<sup>[15]</sup>以及基于单应阵的编队跟随控制<sup>[16]</sup>等, 更有相关工作将同步定位与地图构建(simultaneous localization and mapping, SLAM)与PBVS进行结合实现控制<sup>[22]</sup>.

### 2.1.2 基于图像的视觉伺服

IBVS的基本框架如图3所示, 该方法和PBVS相比无需估计位姿. 将PBVS中的影响精度的位姿估计问题转换为像素空间的控制问题, 在2D图像空间构造成本函数, 通过最小化成本函数来控制当前图像到达期望图像. 由于误差定义在2D图像空间中, 因此也称2D视觉伺服.

IBVS由于是在2D图像空间构建闭环控制, 因此能够直观地确保参考物体在相机的视场内即满足导引头视野范围(field of view, FOV)约束. 直接基于图像空间进行控制, 能降低相机标定误差和机器人的模型误差的影响, 也因此具有很高的稳态控制精度, 能够最大程度地减少图像空间中的定位误差, 即使对于动态定位问题, 例如动态目标跟踪问题, 其准确度依然

很高,因此IBVS的应用也十分广泛,包括在移动机器人上,实现路径跟随的控制<sup>[17]</sup>;在低轨对地观测卫星上,跟踪任何静止或不动的地物<sup>[18]</sup>;以及在解耦控制的基础上结合IBVS实现控制<sup>[23]</sup>.

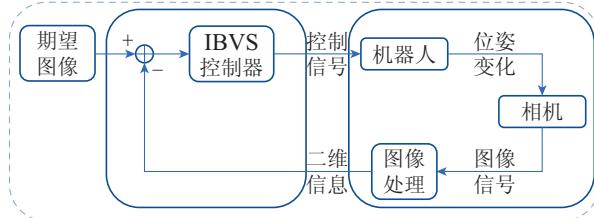


图3 基于图像的视觉伺服

Fig. 3 Image-based visual servo (IBVS)

但IBVS控制系统对于大幅度旋转的伺服控制效果较差,且该方法只具备局部渐近稳定性,无法和PBVS一样在整个任务空间都是收敛的,存在一些其他的问题,如雅可比矩阵奇异或陷入局部最小化等.

图4展示了典型的IBVS的计算流程.为了实现最终的伺服控制,该目标的实现主要分为3大部分.如式(1)所示,首先需要选择合适的图像描述方法,结合图像度量方法来构造误差函数.其次根据所选图像描述特征推导估计出式(2)类似结构的交互矩阵 $\mathbf{L}_s$ (interaction matrix),交互矩阵即图像特征变化速率与机器人末端在欧氏空间中速度的雅可比矩阵.最后如式(3)所示,选择合理的控制律实现最佳收敛. IBVS的最终目的是通过最小化成本函数,分析计算出相机速度 $\nu = \{\mathbf{v}, \mathbf{w}\}$ .对此,针对基于特征点的IBVS的计算过程进行一定说明<sup>[2-4]</sup>.

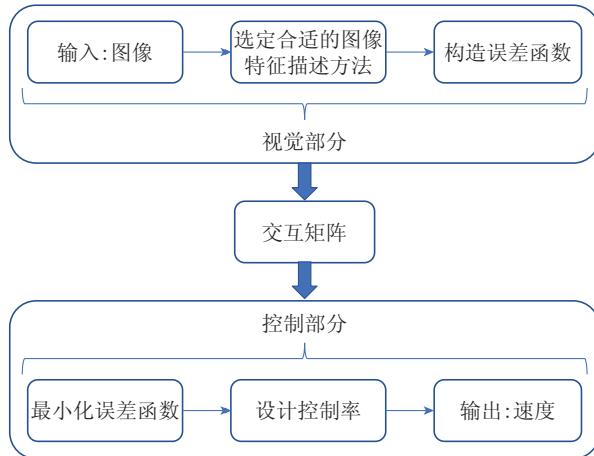


图4 基于图像的视觉伺服的计算流程

Fig. 4 Calculation process of IBVS

1)首先需要选择合适的特征即式(1)中的 $s$ .对于不同的图像特征,所构造的成本函数 $e(t)$ 形式不一,导致最终的交互矩阵推导也有难易之分.典型的IBVS主要利用局部特征以及全局特征两类,使用全局特征

的方法也称为直接视觉伺服<sup>[10-12]</sup>.

$$\begin{cases} e(t) = s - s^*, \\ \dot{s} = \mathbf{L}_s v_c. \end{cases} \quad (1)$$

2)其次需要针对特征求解交互矩阵即式(1)中的 $\mathbf{L}_s$ .基本的以像素坐标为局部特征的IBVS,其交互矩阵即式(2)所示的 $\mathbf{L}_x$ ,其中 $(x, y)$ 为归一化投影坐标.以往的视觉伺服研究主要在于交互矩阵的构造与估计,大多数典型的视觉伺服,由于特征维度较低,因此估计交互矩阵的方法较简单,如使用多项式进行拟合并进行一定的时延补偿<sup>[24]</sup>,而由于DVS使用高维度全局特征,往往需要使用神经网络对交互矩阵中的大量参数进行估计<sup>[25]</sup>.

$$\mathbf{L}_x = \begin{pmatrix} -1/Z & 0 & x/Z & xy & -(1+x^2) & y \\ 0 & -1/Z & y/Z & 1+y^2 & -xy & -x \end{pmatrix}. \quad (2)$$

3)最后设计控制律求出速度即式(3)中 $\nu$ .控制律的设计需要结合成本函数的形式进行,但在最小化成本函数 $e(t)$ 的过程中,IBVS的收敛域与收敛速度往往存在冲突,因此需要进行合理设计,结合视觉伺服任务计算出相应的相机速度.

$$\begin{cases} \hat{r} = \arg \min(s(r) - s^*)^T (s(r) - s^*), \\ \mathbf{v} = -\lambda \hat{\mathbf{L}}_s^+ (s - s^*). \end{cases} \quad (3)$$

### 2.1.3 混合视觉伺服

典型的HVS大多采用Malis的思想,其基本框架如图5所示<sup>[26]</sup>,其核心思路在于分解当前图像与期望图像之间的单应矩阵,获取当前位姿与期望位姿的误差,在3D空间内进行旋转控制,同时利用扩展的图像坐标在2D图像空间内进行位置控制. HVS将部分控制误差定义在2D图像空间中,同时还在3D笛卡尔空间表示出另一部分的误差,因此也称2.5D视觉伺服. HVS结构特点表明该方法可以很好地综合PBVS和IBVS的优点,在保证精度的同时提高系统的稳定性和收敛域. HVS仍然存在一些问题,例如不能确保参考对象满足FOV约束、在解耦单应矩阵时存在不唯一解.

HVS作为一种综合PBVS和IBVS优势的改进方法,其研究一直在进行.在移动机器人领域,Collewet等<sup>[19]</sup>就曾尝试针对复杂的物体或形状未知的物体,使用户HVS的思想进行视觉伺服,该方法在期望图像无法精准描述物体的情况下,假设物体是平面的且静止的,随后基于2D视觉伺服进行控制,并在此基础上估计两个3D参数,最终实现复杂物体的最终形状估计.此外,针对机械臂相关的应用,Hafez等<sup>[20]</sup>通过获取模型参数或进行部分重建,最终实现分解单应矩阵获取完整的2D和3D信息,并以HVS的方式考虑进成本函数内,该方法有效避免相机撤退等问题.此外,对于具有强

耦合特性的机器人如旋翼飞行机械臂, 使用HVS的方式能够实现有效解耦控制, 王耀南等<sup>[21]</sup>通过运动学与动力学联合建模, 实现HVS解耦控制, 改善了系统对

非结构因素的抗扰性能和全局稳定性, 针对非完整移动机器人与机械臂联合的机器人, 通过推导全局雅可比矩阵也可以实现HVS控制<sup>[27]</sup>.

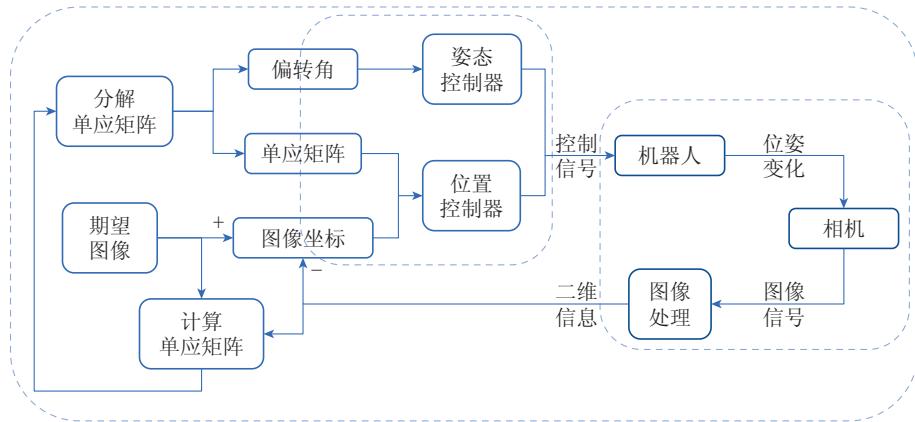


图 5 混合视觉伺服

Fig. 5 Hybrid visual servo (HVS)

## 2.2 基于全局特征的直接视觉伺服

视觉系统的关键在于如何合理利用蕴含的图像信息, 以往的IBVS大多基于图像特征来实现控制任务, 因为移动图像上特征点的坐标暗含位姿的变化<sup>[9-10]</sup>, 且主要利用的图像特征大多是点、线、面甚至是3D姿势或单应性等几何信息, 但利用这些信息不仅要求相机姿态准确, 还要求特征提取够稳健、帧间匹配够快速以及时空跟踪够实时.

此外, 该图像处理过程不仅复杂, 且往往存在计算效率不佳、易受噪声干扰、发生尺度和旋转变化时精度不高、手工提取的特征容易出错且通用性有限等问题. 同时, 基于点特征的交互矩阵具有强烈非线性和强耦合性, 存在收敛域有限、容易陷入局部最小值、容易出现奇异解等问题. 但在基于局部特征视觉伺服中, 这一系列的处理过程却仍被认为是必要的, 也因此成为限制视觉伺服发展的瓶颈之一.

为了规避复杂的图像特征提取流程, 直接视觉伺服(DVS)开始出现<sup>[10-12]</sup>. 该方法跳过了特征抽取的过程, 直接使用全局图像特征如光度、色彩、梯度、深度等, 构造误差函数来设计控制器. 由此可见, 直接视觉伺服中包含两层含义: 首先DVS直接在信号级别进行操作, 不需要进行任何中间估计环节, 包括不限于位姿估计以及特征点提取等环节; 其次, 在控制方面, DVS只使用非度量信息进行控制, 不需要进行任何分解和先验过程.

相较而言, DVS能有效利用部分不存在特征的区域, 避免了图像特征进出视野导致的不连续性与奇异性, 甚至避免了特征的丢失的问题, 有效维护了系统的稳定性. 同时, DVS利用更多的图像信息来描述任

务, 增强了控制系统的准确性和鲁棒性, 可以有效降低定位误差. 因此, 在稳定性、准确性和鲁棒性方面, 基于全局特征的DVS均优于基于局部特征的方法. 但考虑到DVS仍然存在的收敛域较小、可能陷入局部极小值等问题, 后续将结合视觉伺服的架构从视觉以及控制两个部分介绍DVS.

## 3 直接视觉伺服中图像特征描述方法

视觉系统作为直接视觉伺服架构中的一大组成部分具有重大意义, 选择图像特征描述方法作为视觉系统的核心任务, 其优劣不仅影响视觉效果的好坏, 还会影响其他部分如交互矩阵的计算难度、控制器设计的难度、甚至最终伺服结果的优劣.

当前直接视觉伺服中使用的图像特征的描述方法多种多样, 但大多基于几类基本的描述方法如色彩、深度、梯度、光度, 其余大部分则组合相应的特征变换如高斯混合、正交变换等得到. 对此, 对于特征描述方法的介绍, 本文将从基本的特征描述方法, 以及基于光度特征变换的特征描述方法两方面展开, 此外还将介绍度量特征描述方法相似性的部分方法.

### 3.1 图像特征基本描述方法

直接视觉伺服作为一种以全局图像特征为反馈信号的闭环控制方法, 鉴于其在稳定性、准确性、鲁棒性的优异表现, 针对不同的视觉伺服任务, 可以选择不同的图像特征. 在此, 针对能直接从图像获取到的基本的全局特征描述方法, 首先重点介绍了光度特征, 随后考虑到纯光度特征对光照敏感的问题, 还介绍了色彩、深度、梯度对光照不敏感的图像特征描述方法.

表2<sup>[28-42]</sup>展示了上述图像特征描述方法的特点及对应的交互矩阵公式.

表2 图像特征的基本描述符

Table 2 The basic descriptor of the image feature

图像特征的基本描述符	特点	特征 $s(r)$ 选择与交互矩阵 $L$ 相关公式
光度特征 <sup>[28-33]</sup>	避免了任何复杂的图像处理 只用计算梯度, 计算高效且简洁 误差小且收敛精度高 成本函数具有高度非线性 收敛域较小, 受光照变化影响大	$s(r) = I(r) = (I_{1\bullet}, I_{2\bullet}, \dots, I_{N\bullet})$ , $L_I = -(\nabla I_x L_x + \nabla I_y L_y)$
色彩特征 <sup>[34-37]</sup>	避免了任何复杂的图像处理 色彩属性比光度更具区别性 色彩特征的多元化便于先验 受光照变化影响小	$s(r) = L_i^\varphi = \frac{\varphi}{N_i}$ , $\varphi \in \{R, G, B\}$ , $L_{L_i^\varphi} = -\frac{1}{N_i} (\nabla \varphi - L_i^\varphi \nabla N_i)^T L_x$
深度特征 <sup>[38-40]</sup>	避免了任何复杂的图像处理 解决单目相机中深度缺乏的问题 不容易受光照变化影响	$s(r) = Z$ , $L_Z = L_{P_Z} v = L_{P_Z} - A L_x - B L_y$ , $L_{P_Z} = [0 \ 0 \ -1 \ -yZ \ xZ \ 0]$
梯度特征 <sup>[41-42]</sup>	避免了任何复杂的图像处理 有助于使用图像边缘信息即轮廓信息 交互矩阵推导以及成本函数的最小化更复杂 受光照变化影响小	$s(r) = \ \nabla I(r)\ ^2 = \nabla I_x^2(r) + \nabla I_y^2(r)$ , $L_G = 2(\nabla I_x \frac{\partial \nabla I_x}{\partial x} + \nabla I_y \frac{\partial \nabla I_y}{\partial x}) L_x +$ $2(\nabla I_x \frac{\partial \nabla I_x}{\partial y} + \nabla I_y \frac{\partial \nabla I_y}{\partial y}) L_y$

### 3.1.1 纯光度特征

早期的DVS技术大多选择光度作为全局图像特征, 通过比较当前图像和期望图像的所有像素的纯光度进行视觉伺服, 这种方法也因此被称为纯光度视觉伺服(photometric visual servo, PVS)<sup>[28-29]</sup>.

基于光度特征的视觉伺服任务同DVS一样, 需要最小化误差函数来设计控制律. 如式(4)所示, 纯光度视觉伺服把每个像素点的光度都作为视觉特征, 可以选择每个像素强度的平方差的总和作为相似性度量方法时<sup>[28]</sup>, 即

$$\begin{cases} s(r) = I(r) = (I_{1\bullet}, I_{2\bullet}, \dots, I_{N\bullet}), \\ \hat{r} = \arg \min(I(r) - I^*)^T (I(r) - I^*). \end{cases} \quad (4)$$

交互矩阵的推导与估计往往需要结合特征的变化重新计算. 光度作为图像的基本描述方法, 其交互矩阵 $L_I$ , 可用于其余基于光度变换的特征描述方法的交互矩阵推导.

$L_I$ 的推导需要结合一阶泰勒级数展开式, 并基于光度不变性假设以及光流约束方程(optical flow constraint equation, OPCE)进行, 如式(5)所示, 其中光流约束方程将光度随时间的变化与给定像素处的图像运动联系起来.

$$\begin{cases} I(x + dx, t + dt) = I(x, t), \\ \dot{I} = -\nabla I^T L_x v. \end{cases} \quad (5)$$

最终交互矩阵 $L_I$ 如式(6)所示, 其中 $L_x$ 如式(2)所示, 正是在第2节中介绍的基于特征点的IBVS所用到

的交互矩阵, 即

$$L_I = -(\nabla I_x L_x + \nabla I_y L_y). \quad (6)$$

在计算效率方面, 观察交互矩阵 $L_I$ 可以发现, PVS的图像处理过程仅需计算图像空间梯度, 高效且简洁; 在准确性方面, 由于图像的测量只有像素强度, 图像处理这一过程的误差很小, 而且纯光度法直接对像素进行比较, 即使在近似深度、部分遮挡、镜面和低纹理的环境中, 也能实现高精度的控制. 但是, 在收敛性方面, 由于所选的成本函数具有高度非线性, 导致收敛域较小, 要求伺服任务前后图像差异不能太大; 同时特征空间和状态空间之间的高度非线性也限制了收敛域, 无法获得足够的机器人轨迹; 在抗干扰能力上, 由于PVS易受光照变化的影响, 因此包含纹理信息的图片需要稳定的光照条件.

针对PVS光照敏感的问题, 进行了一系列改进. 例如使用模板照明适应的方式, 将参考图像替换为用概率期望运算符计算后的期望图像, 动态地使期望图像适应相机获取当前图像的光照条件<sup>[30]</sup>. Silveira等<sup>[31]</sup>曾利用全反射模型来应对复杂的照明变化, 提高直接视觉跟踪过程中对于光照鲁棒性. Collewet等<sup>[32]</sup>曾针对OPCE光流约束提出了一个新的光流约束, 通过修改时间恒定性的基本假设, 来应对包括镜面高光在内的复杂的亮度变化. 此外, PVS的应用也不限于此. Silveira等<sup>[33]</sup>基于光度设计的非度量视觉伺服技术, 实现平移和旋转控制误差的完全解耦.

### 3.1.2 色彩特征

色彩特征在过去一直受到关注。例如使用色彩特征来实现三维场景的分割<sup>[34]</sup>, 使用彩色图像进行3D检查与物体识别, 此外, 还出现了针对复杂场景以及光照变化情况下的图像分割算法<sup>[35]</sup>。

基于色彩特征的DVS在某种程度上具有同纯光度的DVS一样的优点, 能避免复杂的图像处理, 此外, 能在一定程度上减少对光照变化的影响。同时, 色彩特征的多元化提高了视觉伺服的灵活性。Collewet等<sup>[36]</sup>曾基于R, G, B分量图像进行控制律的设计, 并提出一个基于交互矩阵的条件数的标准, 利用色彩特征的多元化, 实现先验地选择潜在的最佳色彩表示方式, 极大提高了控制律收敛性, 改善了相机轨迹, 且该方式的交互矩阵的计算简单, 只需堆叠相应色彩特征的交互矩阵即可。

基于色彩特征的DVS在控制律的设计方面的研究进展也很亮眼。Silveira等<sup>[37]</sup>针对基于色彩特征的DVS, 提出新的直接转换模型和新的高效的优化方法, 以应对物体变形以及照明变化带来的影响, 达到准确和稳健视觉伺服跟踪目的, 该方法实现了灰度和彩色图像、一般光照变化下、刚性和可变形表面的统一直接视觉跟踪。

式(7)中 $L_i^\varphi$ 作为伺服特征, 可以选择RGB的多种组合以及变式 $N_i$ <sup>[36]</sup>。

$$\begin{cases} s(r) = L_i^\varphi = \frac{\varphi}{N_i}, \varphi \in \{R, G, B\}, i = \{1, 2\}, \\ N_1 = R + G + B, \\ N_2 = \sqrt{R^2 + G^2 + B^2}, \\ N_3 = (R - G)^2 + (R - B)^2 + (G - B)^2. \end{cases} \quad (7)$$

通过推导证明, 可以计算得出最终的交互矩阵 $\mathbf{L}_{L_i^\varphi}$ 如式(8)所示, 相较于光度视觉伺服的交互矩阵式(6), 计算量更大。

$$\mathbf{L}_{L_i^\varphi} = -\frac{1}{N_i} (\nabla \varphi - L_i^\varphi \nabla N_i)^T \mathbf{L}_x. \quad (8)$$

### 3.1.3 深度特征

鉴于单目视觉传感器存在所测图像易受光照影响的问题, 导致PVS在一些情况下失效。对此, RGB-D相机的出现为DVS的发展提供了一种研究思路, 即选择从距离传感器获得的密集深度图作为特征进行视觉伺服。此外, 基于深度特征的DVS能轻松解决单目相机中存在的深度信息缺乏的问题。

深度测量的方式有很多。当前能提供深度信息的非接触式距离传感器大多数是有源的, 如激光扫描、声纳、雷达、ToF或RGB-D相机等, 且大多基于飞行时间(time of flight, ToF)原理进行, 但当已知发送信号速度时, 也可使用调制信号并测量相移来获取深度信息, 或使用结构光来测量深度。此外, 还可通过无源传

感器如摄像机获取深度信息, 在校准过的平台上通过匹配图像特性, 从几何角度计算深度, 虽然深度信息稀疏, 但是也可以获得密集的深度地图, 或者利用图像散焦信息确定目标距离(depth-from-focus), Surya对此也提出了一种新的空间域卷积/反卷积变换来实现更高效的深度计算<sup>[38]</sup>。

当前大多数DVS方法获取的深度信息基于有源传感器, Teuliere等<sup>[39-40]</sup>直接使用距离传感器获取的密集的深度图作为视觉特征, 将定位任务转换为将深度特征调节为期望值的深度特征。该方法不进行3D姿势估计, 避免了任何复杂的图像处理, 即使在场景发生部分遮挡或被修改、场景是非平面的、存在测量误差的情况下, 仍可以正确完成定位任务。在一些参考深度图难以获取的情况下, 即使使用渲染引擎生成的合成深度图, 仍可带来令人满意的定位性能。若使用有源传感器获取深度图, 所得方法可以避免光照变化的影响。虽然使用有源传感器进行深度测量可能会出现一些限制, 尤其是存在噪声和缺乏测量结果的情况下, 但可以通过使用估计器并进行基本的图像预处理解决上述问题。

选择深度 $s(r) = Z$ 作为特征, 计算得出最终的交互矩阵 $\mathbf{L}_Z$ 如式(9)所示, 从表达式可以看到相较于光度视觉伺服, 该方法需要计算一个额外的交互矩阵 $\mathbf{L}_{P_Z}$ , 计算不够高效。

$$\begin{cases} \mathbf{L}_Z = \mathbf{L}_{P_Z} \mathbf{v} = \mathbf{L}_{P_Z} - A \mathbf{L}_x - B \mathbf{L}_y, \\ \mathbf{L}_{P_Z} = [0 \ 0 \ -1 \ -yZ \ xZ \ 0]. \end{cases} \quad (9)$$

### 3.1.4 梯度特征

鉴于光度在发生大非仿射照明变化时存在不稳定性, 尝试使用图像梯度特征。基于图像梯度的方法常用于确定强度快速变化的区域, 鉴于图像中的大多数有用信息位于其高频区域, 因此基于图像梯度的DVS对于图像边缘信息即轮廓信息的提取非常有帮助。其次, 基于图像梯度的DVS对光照变化比PVS更稳健, 也能保证一定的定位误差, 其主要困难是推导与此相关的交互矩阵, 而且选择梯度特征的成本函数的最小化更为复杂。当选择图像梯度作为全局图像特征进行DVS时, 控制目标为最小化当前图像梯度和所需图像梯度的误差<sup>[41]</sup>。

当前基于梯度信息构建控制律的一种方案是使用图像梯度的范数。此外还有一些方法将光度和梯度两种全局特征进行各种组合。此外, 为了解决单目视觉伺服存在的缺乏场景深度的问题, 以及更准确地进行高倍率下的微定位视觉伺服, 需要提高z轴的控制精度, 由于沿z轴运动能引起图像清晰度的变化, 因此使用梯度特征能够准确显现出z轴上的平移变化。基于这一思想的基于离焦深度(depth from defocus, DFD)的方法也被提出<sup>[42]</sup>。这些方法的主要思想是, 距镜头

特定距离的物体将在光学系统中聚焦,而其他距离的物体将变得模糊。通过测量观察到的图像上物体的散焦量,可以从几何光学器件中恢复物体相对于透镜的深度。

当选择式(10)中的梯度范数 $\|\nabla \mathbf{I}(r)\|^2$ 作为视觉伺服的特征时,

$$\mathbf{s}(r) = \|\nabla \mathbf{I}(r)\|^2 = \nabla \mathbf{I}_x^2(r) + \nabla \mathbf{I}_y^2(r). \quad (10)$$

得到的交互矩阵 $\mathbf{L}_G$ 如下:

$$\begin{aligned} \mathbf{L}_G = & 2\left[\left(\nabla \mathbf{I}_x \frac{\partial \nabla \mathbf{I}_x}{\partial x} + \nabla \mathbf{I}_y \frac{\partial \nabla \mathbf{I}_y}{\partial x}\right) \mathbf{L}_x + \right. \\ & \left. \left(\nabla \mathbf{I}_x \frac{\partial \nabla \mathbf{I}_x}{\partial y} + \nabla \mathbf{I}_y \frac{\partial \nabla \mathbf{I}_y}{\partial y}\right) \mathbf{L}_y\right]. \quad (11) \end{aligned}$$

### 3.2 基于光度变换的特征描述方法

现阶段的DVS大多基于光度变换的特征描述方法,同时考虑到纯光度DVS的窄收敛域往往是由于成本函数的高度非线性造成的,而基于光度变换的DVS不再考虑空间域中的图像本身,而是通过多种方法进行特征空间的转换,在新的特征空间中进行DVS。这类本身具备降维能力的光度变换方法,在保留大部分信息的基础上压缩原始图像信息,既可以保证原有的准确度也可以降低成本函数的非线性。理论上讲,只要问题的维度越低,成本函数更平滑,收敛区域更大,成本函数梯度更高,收敛速度将更快。

通过对这类基于光度变换的特征描述方法分类整理发现,主要包含直方图特征表示方法、矩特征表示方法、高斯混合特征表示方法、散焦特征表示方法、基于正交变换的频域空间表示方法、基于压缩反馈的集合空间表示方法、基于自动编码的潜在空间表示方法这几种。表3<sup>[43-71]</sup>展示了上述基于光度变换的特征描述方法的特点及对应的交互矩阵公式。

#### 3.2.1 直方图特征

直方图作为一种仅用一组值便可实现全局描述的强大的方式,在计算机视觉领域中有着应用。无论是使用颜色直方图作为描述符的Mean Shift算法,还是基于定向梯度直方图计算的Sift特征,都说明了使用直方图来进行视觉伺服的通用性。此外,直方图作为一种估计变量的概率分布的方式,其背后的主要思想是根据该像素的值将其分到相应的类中。每个类由一个标量值组成,该标量值每次找到相应的元素时都会递增,并且一旦测试了图像的每个像素,类的值将根据图像中的像素数量进行归一化<sup>[43]</sup>。

使用灰度强度直方图作为视觉特征时,可以通过最小化两个直方图之的马氏距离(Jeffries Makushita)来进行DVS<sup>[44]</sup>。对整个图像或多个子集中使用强度直方图这一全局特征,可以更紧凑地表示图像信息,对噪声的鲁棒性增强,同时也可以通过选择直方图中类的数量来提高伺服算法灵活性。受基于核方法的启

发<sup>[72-73]</sup>,在直方图的计算中引入加权核函数,分别为图像中心和边缘赋予不同权重,可以增强对于图像像素的出现与消失的鲁棒性。将图像划分为多个区域并将直方图与每个区域相关联,然后堆叠生成的误差向量和交互矩阵,可实现多自由度控制。但不可避免的是,基于强度直方图的DVS仍然会受光照变化的严重影响,由此出现了颜色直方图。

颜色直方图不仅具有强度直方图的上述优点,还能获取此外更多信息。为了减少光照变化对基于强度直方图的DVS的影响,可以通过整合颜色信息进行。一种是只关注颜色信息,为此需要将颜色空间从RGB变到HSV,实现色相和饱和度的颜色信息与纯强度分离。另一种是通过选择HSV原始图像的两个颜色通道并求和合成强度图像,也可降低对光照变化的敏感。

定向梯度直方图与强度直方图的不同之处在于,它不直接对原始图像像素的值进行分类,而是依赖于每个像素的梯度方向的计算<sup>[43]</sup>。在提高对光照变化的鲁棒性方面,因为梯度方向对于全局或局部的光照变化是不变的,梯度特征十分有效。此外,与纯光度DVS进行纯像素级别的比较相反,直方图可用于比较信息量有限的图像;信息量的减少不仅可能会提高局部变化的鲁棒性,此外,可以使用多种值来计算直方图进行灵活比较。例如针对不同光照变化,通过采用每个像素的梯度方向而不是其强度值。

式(12)选择单纯的灰度强度直方图 $\mathbf{p}_I(i)$ 作为伺服特征,并选择马氏距离作为衡量两个直方图之间的成本函数 $\rho(\mathbf{I}, \mathbf{I}^*)$ 。

$$\begin{cases} \mathbf{p}_I(i) = \frac{1}{N_x} \sum_x^{N_x} \delta(\mathbf{I}(x) - i), \\ \rho(\mathbf{I}, \mathbf{I}^*) = \sum_i^{N_c} (\sqrt{\mathbf{p}_I(i)} - \sqrt{\mathbf{p}_I^*(i)})^2, \\ \mathbf{L}_H = \frac{\partial \rho(\mathbf{I}, \mathbf{I}^*)}{\partial r}. \end{cases} \quad (12)$$

获得的相应的交互矩阵 $\mathbf{L}_H$ 如式(13)所示:

$$\mathbf{L}_H = N_c \sum_i^{N_c} \left( \frac{\partial \mathbf{p}_I(i)}{\partial r} \left( 1 - \frac{\sqrt{\mathbf{p}_I^*(i)}}{\sqrt{\mathbf{p}_I(i)}} \right) \right). \quad (13)$$

#### 3.2.2 矩特征

在控制6个自由度的视觉伺服任务中,或当前图像与期望图像误差较大时,交互矩阵往往具有高度非线性和强耦合性质,此时的收敛域有限解,且容易陷入局部最小值,出现奇异解类的问题。

为了解决由高度非线性和强耦合带来的上述问题,尝试在视觉伺服中引入矩特征(moment)。在基于局部特征的IBVS中,使用基于几何矩的方法<sup>[45-46]</sup>极大降低了系统的非线性,增大了IBVS的收敛域也提高了机器人轨迹的合理性。

上述几何矩是基于一组图像点构建的,借鉴几何

矩的思路, 尝试在DVS的背景下, 基于像素光度值构建一组光度矩特征来进行视觉伺服<sup>[47-48]</sup>。光度矩是基于核方法的特定情况<sup>[72-73]</sup>, 但与之不同的是, 该方法可以考虑所有6个自由度, 还可以以封闭形式推导

交互矩阵, 研究表明此时交互矩阵的结构呈现出良好的解耦特性。然而, 由于交互矩阵的建模是在恒定的图像边界假设下进行的, 当场景的一部分进入或离开摄像机的视场时, 会导致失败。

表3 基于光度变换的特征描述方法

Table 3 Feature description method based on photometric transformation

特征描述方法	特点	特征 $\mathbf{s}(r)$ 与对应的交互矩阵 $\mathbf{L}$
直方图特征 <sup>[43-44]</sup>	密集的特征增强了对噪声、遮挡的鲁棒性 可选方式多, 算法更灵活性 颜色、梯度直方图对光照不敏感	$\mathbf{s}(r) = \mathbf{p}_I(i) = \frac{1}{N_x} \sum_x^{N_x} \delta(I(x) - i),$ $\rho(\mathbf{I}, \mathbf{I}^*) = \sum_i^{N_c} (\sqrt{\mathbf{p}_I(i)} - \sqrt{\mathbf{p}_{I^*}(i)})^2,$ $\mathbf{L}_H = N_c \sum_i^{N_c} \left( \frac{\partial \mathbf{p}_I(i)}{\partial r} \left( 1 - \frac{\sqrt{\mathbf{p}_I^*(i)}}{\sqrt{\mathbf{p}_I(i)}} \right) \right)$
矩特征 <sup>[45-51]</sup>	良好的解耦特性降低了系统的非线性 收敛域增大收敛速度快 避免陷入局部最小值 提高了机器人轨迹的合理性	$\mathbf{s}(r) = m_{pq} = \iint_{\pi} x^p y^q w(x) I(x, t) dx dy,$ $\mathbf{L}_{m_{pq}} = \iint_{\pi} x^p y^q w(x) \mathbf{L}_I dx dy$
高斯特征 <sup>[52-55]</sup>	收敛域更大 可利用高斯参数调节更灵活 算法更复杂且更耗时间	$\mathbf{s}(r) = G_1 = \sum_{u \in U} I(u) \exp\left(-\frac{(u_g - u)^2 + (v_g - v)^2}{2\lambda_g^2}\right),$ $\mathbf{s}(r) = G_2 = \sum_{u \in U} I(u) \exp\left(-\frac{(u_g - u)^2 + (v_g - v_i)^2}{2I^2(u)\lambda_g^2}\right),$ $\mathbf{L}_G = -\sum_u (\nabla \mathbf{I}^T E(u_g - u) \mathbf{L}_u)$ $\mathbf{s}(r) = I_d(u) = \int_X I(pr(X)) \tilde{g}(u, X) dX,$ $\tilde{g}(u, X) = \frac{1}{2\pi\lambda(Z)^2} \exp\left(-\frac{\ u - pr(X)\ ^2}{2\lambda(Z)^2}\right),$ $\mathbf{L}_{I_d}(u) = \begin{bmatrix} -\nabla_u \mathbf{I}_d^T \\ -\Delta_u \mathbf{I}_d \end{bmatrix} \begin{bmatrix} \mathbf{L}_u \\ \frac{Df}{6k_u(Z_f - f)Z} \mathbf{L}_Z \end{bmatrix}$ $\mathbf{s}(r) = w_k(r) = \mathbf{U} k^T (\mathbf{I} - \bar{\mathbf{I}}),$ $\mathbf{I}_R = \bar{\mathbf{I}} + \sum_{k=1}^K w_k \mathbf{U}_k^T,$ $\mathbf{L}_{w_k} = \mathbf{U}_k^T \mathbf{L}_I$
散焦特征 <sup>[56-57]</sup>	大范围收敛性能类同高斯特征 算法更简单消耗时间短	$\mathbf{s}(r) = \mathbf{f}(r) = \text{vec}(\mathbf{F}(r)),$ $\mathbf{F}(r) = \mathbf{C} \mathbf{I}(r) \mathbf{C}^T,$ $\mathbf{L}_F = \mathbf{C} \frac{\partial \mathbf{I}(r)}{\partial r_i} \mathbf{C}^T$ $\mathbf{s}(t) = (I(u_1, v_1, t) \cdots I(u_h, v_w, t))^T,$ $\mathbf{L}_{I(u,v,t)} = (\nabla y \ t(u, v))^T \mathbf{L}_p(u, v) \mu(t),$ $\mathbf{L}_p = \begin{bmatrix} 1/z & 0 & u/z & uv & (1+u^2) & v \\ 0 & 1/z & v/z & 1+v^2 & uv & u \end{bmatrix}$
基于KL变换的频域空间表示 <sup>[58-60]</sup>	频域空间使用PCA分解降维更简单 成本函数更平滑, 收敛区域更大	$\mathbf{s}(r) = \mathbf{z} = \mathbf{W} \mathbf{I} + \mathbf{b},$ $\mathbf{L}_z = \frac{\partial \mathbf{W} \mathbf{I} + \mathbf{b}}{\partial r} = \mathbf{W} \frac{\partial \mathbf{I}}{\partial r} + \frac{\partial \mathbf{W}}{\partial r} \mathbf{I} = \mathbf{W} \mathbf{L}_I$
基于DCT变换的频域空间表示 <sup>[61-67]</sup>	频域空间挑选低频系数降维更简单 成本函数更平滑, 收敛区域更大	
基于压缩反馈的集合空间表示 <sup>[68-70]</sup>	在线性结构不适用的集合空间中设计 采样和压缩同时进行 计算负担小, 重构能力强 获取压缩反馈的方式灵活	
基于自动编码的潜在空间表示 <sup>[71]</sup>	无监督学习潜在的转换空间 数据表示更丰富且合理 压缩维度更小且重构能力更好 成本函数更平滑, 收敛域更大, 精度更好	

DVS中一个不可忽视的图像变化, 即新场景的出现或旧场景从相机视场中消失。这类未建模的干扰通常可能导致控制律失效。对此, 提出了一种通过空间加权增强的光度矩的通用模型, 给出了这些空间加权光度矩的交互矩阵的解析形式<sup>[48]</sup>。确定一个加权函

数, 然后利用基于矩不变量的现有结果从光度矩中获得视觉特征, 使用这些视觉特征的伺服控制适用于当前图像与期望图像重叠较少的情况。与纯光度特征和基于非加权矩的特征相比, 具有较大的收敛域, 而且该方法可以用于非平面情况, 也能为实现小位移提供

了令人满意的结果<sup>[49–51, 74]</sup>.

光度矩 $m_{pq}$ 的表达式如式(14)所示:

$$s = m_{pq} = \iint_{\pi} x^p y^q w(x) I(x, t) dx dy. \quad (14)$$

当选择光度矩作为特征时, 相应的交互矩阵 $\mathbf{L}_{m_{pq}}$ 如式(15)所示:

$$\mathbf{L}_{m_{pq}} = \iint_{\pi} x^p y^q w(x) \mathbf{L}_I dx dy. \quad (15)$$

### 3.2.3 高斯混合特征

在经典的IBVS中, 图像点特征被统一建模为高斯混合物, 视觉伺服通过最小化当前和所需姿势下高斯混合之间的距离函数进行<sup>[52]</sup>. 将点的几何特征表示为高斯分布的想法, 可以规避点的匹配和跟踪, 但仍需要进行特征点检测, 因为该方法要求当前图像中提取的点与从期望图像中提取的点完全相同.

将基于点的高斯混合建模引申到所有图像像素点的光度特征建模, 便可构建光度高斯混合模型(photometric gaussian model, PGM)<sup>[53–55]</sup>. 基于光度高斯特征的交互矩阵建模可以采用多种方式: 其中基于格林定理(Green's theorem modeling, GTM)的方法使用格林定理来规避交互矩阵中图像梯度的计算, 此外, 基于高斯光度一致性(photometric Gaussian consistency, PGC)的方法则将三维场景本身视作一个高斯混合物, 图像是三维场景在二维平面上的投影结果, 进一步简化了交互矩阵的计算.

将每个像素点建模为一个光度高斯混合物, 可以放大DVS的收敛域, 即使期望图像和初始图像几乎没有重叠时, 也能确保收敛, 这对于其他的光度变换特征而言通常是一个挑战. 此外, 即使选择每个像素强度的平方差的总和作为误差函数, 基于PGM的视觉伺服也可以通过高斯扩展参数的调节保证较大的收敛域, 同时还能确保其具有和纯光度PVS一样精确的收敛情况.

当选择高斯混合特征 $G(\mathbf{I}, u_g, \lambda_g)$ 作为伺服特征进行PGM时, 高斯混合方式可以有如式(16)所示的两种, 其区别在于是否将光度考虑进高斯函数内.

$$\begin{cases} G_1 = \sum_{u \in U} I(u) \exp\left(-\frac{(u_g - u)^2 + (v_g - v)^2}{2\lambda_g^2}\right), \\ G_2 = \sum_{u \in U} I(u) \exp\left(-\frac{(u_g - u)^2 + (v_g - v_i)^2}{2I^2(u)\lambda_g^2}\right). \end{cases} \quad (16)$$

得到的交互矩阵 $\mathbf{L}_G$ 如式(17)所示:

$$\begin{cases} \mathbf{L}_G = -\sum_u (\nabla \mathbf{I}^T E(u_g - u) \mathbf{L}_u), \\ \mathbf{L}_u = \begin{bmatrix} \alpha_u & 0 \\ 0 & \alpha_v \end{bmatrix} \mathbf{L}_x. \end{cases} \quad (17)$$

分析结果可发现, 虽然纯PVS收敛结果非常准确, 收敛速度很快, 不需要任何参数调整, 但收敛域是有限的. 而基于PGM的DVS的收敛结果也很准确, 收敛域也远大于基于直方图特征和矩特征的DVS方法, 也更显著性的大于基于纯PVS, 但其交互矩阵的计算复杂度呈平方趋势增加, 耗时严重.

### 3.2.4 散焦特征

散焦属性, 也被称为“波克效应”, 常用来实现逼真的渲染, 其中图像中的散焦模糊量取决于场景点和聚焦平面之间的深度差<sup>[56]</sup>. 在计算机视觉领域, 散焦特征多用于计算深度, 用来导航定位等<sup>[38]</sup>.

散焦模糊方法受到基于PGM的DVS方法的启发, 期望通过散焦模糊实现PGM中高斯模糊图像的效果, 进而改善控制律的收敛域<sup>[57]</sup>. 事实上, PGM是具有各向同性的高斯模糊, 而与散焦有关的模糊是各向异性的, 两者的收敛域是具有一定的相似性, 并且基于散焦属性的DVS的收敛域可能大大超过了最先进的基于PGM的视觉伺服的收敛域. 此外, 该方法使用光度描述特征, 最终收敛效果的准确度也能满足. 最重要的是, 由于散焦图像是可以直接获取, 而基于PGM的高斯模糊是对清晰图像进行处理计算的结果, 相较而言, 该方法能极大程度上缩短处理时间.

散焦特征 $\mathbf{I}_d(u)$ 的表达式如式(18)所示:

$$\begin{cases} \mathbf{I}_d(u) = \int_X \mathbf{I}(pr(X)) \tilde{g}(u, X) dX, \\ \tilde{g}(u, X) = \frac{1}{2\pi\lambda(Z)^2} \exp\left(-d \frac{\|u - pr(X)\|^2}{2\lambda(Z)^2}\right). \end{cases} \quad (18)$$

当选择散焦特征作为伺服特征时, 对应的交互矩阵 $\mathbf{L}_{I_d}(u)$ 如式(19)所示:

$$\mathbf{L}_{I_d}(u) = \begin{bmatrix} -\nabla_u \mathbf{I}_d^T \\ -\Delta_u \mathbf{I}_d \end{bmatrix} \begin{bmatrix} \mathbf{L}_u \\ \frac{Df}{6k_u(Z_f - f)Z} \mathbf{L}_z \end{bmatrix}. \quad (19)$$

### 3.2.5 基于正交变换的频域空间表示方法

由于相邻的像素的光度是高度相关的, 所以图像空间的纯光度PVS使用全部像素这一方法是冗余的, 因此, 为了提高收敛性, 需要进行相关冗余数据的祛除, 但如果在图像空间中进行, 这一选择和匹配过程将无比繁琐. 因此尝试将伺服转换到频域空间进行. 其主要优势是, 可以结合时频域之间变换的特点剔除不必要的系数, 同时在频域进行这一降维操作选择最佳频率是非常简单的, 研究表明合理选择系数不仅可以减少问题的维度, 还能很好地改善控制行为. 因此将直接视觉伺服发展到频域空间进行<sup>[75]</sup>.

频域空间的视觉伺服很少见, 当前主要通过正交变换实现频域空间的转换, 并通过选择参数来实现降维. 其中就包含使用KL(Karhunen-Loeve)正交变换变

换到频域后, 再利用主成分分析(principal component analysis, PCA)降维到子空间内进行伺服, 或者使用傅里叶变换相关的正交变换变换到频域, 然后通过合理选择系数实现图像降维。

KL变换作为一种线性正交变换<sup>[58]</sup>, 其变换矩阵有多种选择, 当KL变换矩阵为协方差矩阵时, 等同于主成分分析PCA, 对于图像而言具有降维效果, 经过PCA变换之后的特征空间称为频域子空间<sup>[59-60]</sup>。基于PCA的DVS不直接使用图像强度进行控制, 而是将图像投影到一个新的正交基上, 然后在此子空间内离线计算图像在该正交基上的坐标, 将其作为新的特征量在子空间进行伺服控制。虽然从能量压缩的角度来看, 该方法实现了特征空间的分解, 降低了图像数据的维数, 成本函数更加平滑, 收敛区域更大。但该方法中选择的正交基函数取自图像数据的统计特性, 必须根据不同场景进行学习, 同时, 与子空间相关的交互矩阵并非通过解析计算推导得出, 而是在离线过程中基于学习的方式估计而来, 增大了伺服任务的工作量。

KL变换中的系数 $w_k$ 的表达式如式(20)所示:

$$\begin{cases} \mathbf{I}_R = \bar{\mathbf{I}} + \sum_{k=1}^K w_k \mathbf{U}_k^T, \\ \mathbf{s} = w_k(r) = \mathbf{U}_k^T (\mathbf{I} - \bar{\mathbf{I}}). \end{cases} \quad (20)$$

当选择KL变换中的系数作为视觉特征时, 其相应的交互矩阵 $\mathbf{L}_{w_k}$ 如式(21)所示:

$$\mathbf{L}_{w_k} = \mathbf{U}_k^T \mathbf{L}_I. \quad (21)$$

以往基于傅立叶变换(Fourier transform, FT)的控制律的研究<sup>[61]</sup>表明了该方法的鲁棒性和准确性, 基于傅立叶变换的方法已经被考虑用来控制沿光轴和围绕光轴的平移和旋转<sup>[62]</sup>。基于傅立叶变换的相关扩展工作, 如离散余弦变换(discrete cosine transform, DCT)、小波变换(wavelets transform, WT)、小剪子波变换(shearlet transform, ST), 在视觉伺服领域正被广泛研究着。

DCT是一种在频域中通过用不同频率振荡的余弦函数之和来表示图像的方式<sup>[63]</sup>, 其中低频区对应图像的连续区域中包含的缓慢变化的信息, 高频区对应图像的边缘区域中包含的快速变化的信息。事实上, 如果考虑到DCT的所有系数, 在频域工作的视觉伺服等价于在空间域的纯PVS, 但因为DCT中的低频区包含了图像的绝大部分信息, 可以将其视为一个低通滤波器, 通过在频域内精心挑选DCT中的少量系数高效完成PVS。通过选择部分系数降低问题维度, 可以保证更平滑的成本函数, 从而提高收敛性。其次, 只考虑低频的部分系数可以增加大尺度结构之间的重叠, 增加收敛域。再者DCT作为一种低通滤波方法, 放弃高频可以抑制图像中的噪声。基于DCT的方法与基于PCA变换到子空间的方法相比只是修改了正交基, 但由于

基函数是预先计算的, 与场景无关, 因此不需要任何学习步骤, 计算速度更快。同时在能量压缩方面, 两者成本函数的形状和收敛区域非常接近, 效果更好。

WT是小波的正交函数序列对信号的表示, 可以在时域和频域中同时产生信号表示, 且两者之间能够实现无损的逆变换<sup>[64-65]</sup>。连续的WT可以看作是傅立叶变换的扩展, 但标准傅立叶变换仅仅能在频域中表示。基于WT的DVS可以通过精心选择数据实现降维, 也可通过多分辨率WT的方式实现, 最终实现增强收敛域的同时保证伺服任务的高精度。该方法的本质是在图像的分解重构过程中, 具有不丢失任何信号细节且以不同分辨率表示图像的能力, 它使用多个分辨率系数来表示图像在空间域中的小波变换。ST中的小剪子波是小波的延伸<sup>[66-67]</sup>。但关于小剪子系数的交互矩阵的计算是数值化的, 并且是离线估计的。与小波类似, 小剪子在去噪和表示各向异性的图像特征方面非常有效, 在稳健性、稳定性、准确性方面也都具有优异表现。

DCT变换中的系数 $f(r)$ 如式(22)所示:

$$\begin{cases} \mathbf{s}(r) = \mathbf{f}(r) = \text{vec}(\mathbf{F}(r)), \\ \mathbf{F}(r) = \mathbf{C}\mathbf{I}(r)\mathbf{C}^T. \end{cases} \quad (22)$$

当选择DCT变换中的系数作为视觉特征时, 其相应的交互矩阵 $\mathbf{L}_F$ 如式(23)所示:

$$\mathbf{L}_F = \mathbf{C} \frac{\partial \mathbf{I}(r)}{\partial r_i} \mathbf{C}^T. \quad (23)$$

### 3.2.6 基于压缩反馈的集合空间表示方法

鉴于大尺寸的图像信号会带来沉重的计算负担, 而对于控制器而言, 庞大的反馈数据也会降低反馈速率。压缩反馈方法的基本思想是用一组集合而不是传统的特征向量来表示反馈图像, 对此, 随之提出另一种改进的PVS方法<sup>[68]</sup>。

该方法与在向量空间中研究的PVS方案根本不同, 所设计的伺服控制器是在向量空间的线性结构不适用的集合空间中, 控制误差和动力学也基于集合空间定义, 控制目标也是使初始压缩反馈集与期望的图像集合误差降至零。和Nyquist采样定理相比, 压缩传感可用更少量的测量重建稀疏或可压缩信号。传统方案中传感器首先以奈奎斯特速率感测大量冗余数据, 然后采用压缩策略丢弃冗余信息以减小信号的大小, 这种方法的采样率和计算成本往往很高, 相比之下, 压缩传感将采样和压缩结合在一起并同时进行, 这在很大程度上减轻了计算负担。

当只进行部分信息的压缩反馈时, 如将一部分像素方向的图像信息用作反馈时, 非向量空间的控制方案仍然有效, 该方法可达到纳米级的运动控制<sup>[69]</sup>。此外, 在非向量空间中提出的压缩反馈控制方案, 也可基于压缩反馈图像推导交互矩阵, 甚至获取压缩反馈

的方式更灵活<sup>[70]</sup>,除了感测矩阵可以为任何随机矩阵外,所提出的方法还允许全局压缩反馈信息。

使用集合空间表示的伺服特征如式(24)所示:

$$\mathbf{s}(t) = (I(u_1, v_1, t) \cdots I(u_h, v_w, t))^T. \quad (24)$$

与之相应的交互矩阵  $\mathbf{L}_{I(u,v,t)}$  如式(25)所示:

$$\begin{cases} \mathbf{L}_{I(u,v,t)} = (\nabla y \ t(u,v))^T \mathbf{L}_p(u,v)\mu(t), \\ \mathbf{L}_p = \begin{bmatrix} 1/z & 0 & u/z & uv & (1+u^2) & v \\ 0 & 1/z & v/z & 1+v^2 & uv & u \end{bmatrix}. \end{cases} \quad (25)$$

### 3.2.7 基于自动编码的潜在空间表示方法

上述方法大多基于现有数学方法来实现特征空间的转换与图像的降维。但如何选择一个合适的空间进行伺服成为了后续问题,因此可选择具有学习能力的网络来实现,学习合适的潜在空间并进行空间的转换。

自动编码器(autoencoder, AE)可用来学习在一些约束条件下如何重建原始输入的网络。AE旨在遵循流形假设基础上,学习数据的良好表示,将输入压缩为维度小得多的数据,且其重构能力比对应的线性产物如PCA更好。故尝试使用AE实现对PVS的改进<sup>[71]</sup>。

该方法基于卷积结构,可以提取出相较于其他降维方法更丰富的特征,由于平滑了成本函数,在精度方面通常与最好的DVS方法相当,且有更好的收敛域。与其他使用深度学习结合视觉伺服的方法相比,该方法以无监督的方式进行训练,学习压缩给定的输入,

能够很好地降低模拟和真实世界之间的差距。重要的是,该方法还能基于网络推导建立学习到的表征和摄像机运动之间的联系,允许机器人在潜在空间中的控制,同时还能使用潜在的表征生成图像类型的子目标,可以作为视觉规划的一种补充。

针对式(26)所示的特定形式的编码器,如果选择输出  $\mathbf{z}$  作为伺服特征,则

$$\mathbf{s}(r) = \mathbf{z} = \mathbf{W}\mathbf{I} + \mathbf{b}. \quad (26)$$

可通过求导计算出式(27)所示的交互矩阵  $\mathbf{L}_z$ ,同样也可以在此基础上对其余网络进行分析。

$$\mathbf{L}_z = \frac{\partial \mathbf{W}\mathbf{I} + \mathbf{b}}{\partial r} = \mathbf{W} \frac{\partial \mathbf{I}}{\partial r} + \frac{\partial \mathbf{W}}{\partial r} \mathbf{I} = \mathbf{W} \mathbf{L}_I. \quad (27)$$

### 3.3 特征的相似性度量方法

视觉伺服任务是为了实现当前图像与期望图像相似性最大化,因此一个完整的伺服系统不仅需要特定的特征,还需要选择合适的相似性度量方法,用相似性度量方法来表示所选特征描述方法之间的差异。选择一个合适的标准对当前图像特征和期望图像特征进行相似性度量,是继特征选择之后影响视觉伺服控制效果的另一关键。

表4<sup>[30, 76-82]</sup>展示了视觉伺服控制中曾使用过的相似性度量手段。图6展示了在  $(t_x, t_y)$  子空间下使用每个像素强度的平方差的总和(sum of squared differences, SSD)、互信息、条件方差的成本函数的大致形状。

表 4 特征的相似性度量方法

Table 4 A measure of similarity of features

相似性度量方法	特点	相似性度量相关公式
经典度量方法 <sup>[76]</sup>	以SSD、马氏距离、互相关为主 计算简单但收敛域小 受光照变化影响大	$SSD = (\mathbf{s}(r) - \mathbf{s}^*)^T(\mathbf{s}(r) - \mathbf{s}^*),$ $Matusita = \sum_i^{N_c} (\sqrt{\mathbf{p}_I(i)} - \sqrt{\mathbf{p}_{I^*}(i)})^2$
互信息度量方法 <sup>[77-81]</sup>	对大范围变化如遮挡、照明变化具有鲁棒性 定位任务准确且能解决多模态对齐问题 计算复杂,计算量大	$MI = -\sum_{i,j} p_{II_k}(i,j) \log\left(\frac{p_{II_k}(i,j)}{p_I(j)p_{I_k}(j)}\right),$ $\mathbf{L}_{MI} = \sum_{i,j} \mathbf{L}_{pII_1}(1 + \log(\frac{p_{II_k}}{p_{I_k}}))$
条件方差总和 <sup>[30, 82]</sup>	具有非线性照明变化的不变性 计算量小能够快速应对复杂场景变化	$SCV = \mathbf{I}(x) - \sum_i i \frac{P_{II^*}(i,j)}{P_{I^*}(j)},$ $\mathbf{L}_{SCV} = \frac{\partial \mathbf{I}(r)}{\partial r} = -\nabla \mathbf{I}(r)^T \mathbf{L}_x$

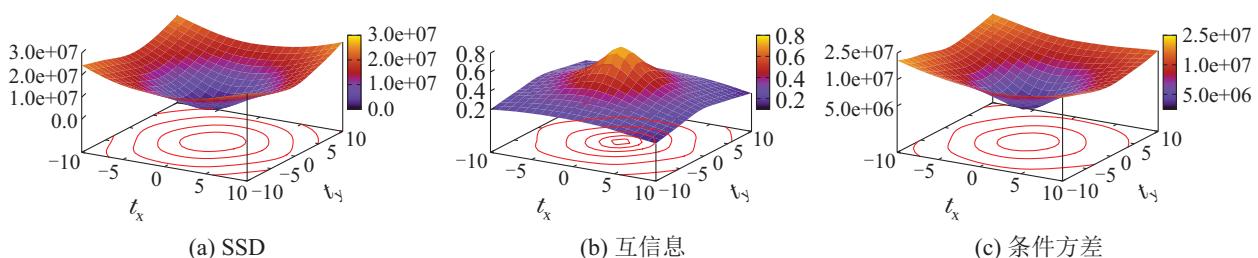


图 6 在  $(t_x, t_y)$  子空间下3种度量方式的成本函数<sup>[82]</sup>

Fig. 6 Pose selection based on particle filtering

### 3.3.1 基于光度的经典度量方法

以往基于光度的度量方法, 直接使用图像所有像素点强度的差值进行设置, 虽然计算量简单, 但往往难达到合适的收敛效果。由此出现了多种经典的相似性度量方法: 多数DVS使用的方法类似于式(4), 即用每个像素强度的平方差的总和(SSD)作为相似性度量方法; 直方图特征使用的方法类同式(12), 即常用两个图像的加权直方图之间的Bhattacharyya距离或Matusita距离进行相似性度量; 也可用归一化互相关(normalized cross correlation, NCC)或零均值归一化互相关(zero mean NCC, ZNCC)的方法进行特征的相似性度量, ZNCC的归一化对线性光度变化非常鲁棒<sup>[76]</sup>。

但对于DVS尤其是PVS而言, 由于图像测量仅是像素强度, 虽然在特征提取过程中几乎没有误差, 但大多数DVS需使用式(5)所示亮度不变性假设, 而随着伺服过程中场景的光的变化, 该假设往往难以满足。因此考虑到图像像素强度对环境变化非常敏感, 为达到精准伺服的目的, 须考虑更稳健的度量方法。

### 3.3.2 基于熵的互信息度量方法

为了提高视觉伺服的鲁棒性, 不直接使用图像的强度, 而是考虑两组像素强度之间共享的信息即互信息。信息论的角度表面, 熵可用来表示图像所包含的信息量, 虽然两个变量的联合熵可以度量整个系统的全局可变性, 但联合熵依赖于两个变量的边际熵, 不适合用作相似性度量。

互信息作为边际熵与联合熵的差, 仅表示两个变量共享的信息量, 更适合作为相似性度量<sup>[77]</sup>。在DVS中, 互信息可用来代表两幅图像共享的视觉信息的数量, 可通过组合当前图像和期望图像之间的互信息和非线性优化过程, 将两者作为配准任务来实现视觉伺服<sup>[78-79]</sup>。

在PVS中使用互信息具有一定优势。首先该方法也不需要任何特征提取与图像处理, 唯一要求是物体具有纹理特征, 由于互信息自然对部分遮挡和照明变化具有鲁棒性, 而控制律只不过是互信息的优化, 因此对场景的外观变化也具有鲁棒性, 这也意味着使用互信息设计伺服控制的方法, 非常适合于大范围的变化如遮挡、镜面反射。准确性方面, 互信息直接考虑图像的全部内容, 且控制过程无需跟踪和匹配过程, 因此最终的定位任务非常准确。互信息的定义指出, 不需要所需图像和当前图像强度之间的线性依赖性, 因此, 即使互信息是从不同的模式中获取的, 只要共享足够的信息, 也能够对齐两个图像, 这对处理多模态对齐问题具有重要意义。但不可避免地是该方法计算相当复杂, 因此计算成本十分昂贵。

图像 $I$ 和 $I_k$ 的互信息 $\text{MI}(I, I_k)$ 如式(28)所示, 计算十分复杂。

$$\left\{ \begin{array}{l} \text{MI}(I, I_k) = - \sum_{i,j} p_{II_k}(i, j) \log\left(\frac{p_{II_k}(i, j)}{p_I(i)p_{I_k}(j)}\right), \\ p_I(i) = \frac{1}{N_x} \sum_x \phi(i - I(x)), \\ p_{II_k}(i, j) = \frac{1}{N_x} \sum_x \phi(i - I(x))\phi(j - I_k(x)). \end{array} \right. \quad (28)$$

此时的伺服控制目标和以往最小化成本函数的方式不同, 需要最大化互信息; 同时由于互信息并非直接基于像素计算信息, 因此交互矩阵 $\mathbf{L}_{\text{MI}}$ 如式(29)所示, 计算十分复杂。

$$\left\{ \begin{array}{l} \rho = \arg \max_{\rho} (\text{MI}(I(\rho), I_k)), \\ \mathbf{L}_{\text{MI}} = \sum_{i,j} \mathbf{L}_{p_{II_1}} (1 + \log\left(\frac{p_{II_k}}{p_{I_k}}\right)). \end{array} \right. \quad (29)$$

对此, 出现了部分工作来降低交互矩阵的复杂性提高伺服效率。有研究表明将车辆运动与互信息之间的交互矩阵推导到二阶, 可以实现更大的收敛域及更快的计算, 同时还创建了一种新的优化方法来处理MI的准凹形, 更好地利用MI的宽收敛域及精确的最大值<sup>[80]</sup>; 此外, 基于Kullback-Leibler散度即相对熵的误差度量方法也可实现更快的收敛速率<sup>[83]</sup>。关于互信息的应用, 在视觉导航的学习阶段储存的自动选择的关键帧, 可以利用当前的图像和附近的关键帧之间的共享信息来实现关键帧的切换<sup>[81]</sup>。

### 3.3.3 基于光度概率分布的条件方差总和的方法

分析上述两种相似性度量方法, 虽然经典度量方法计算量较少, 但对照明的变化不鲁棒, 即使ZNCC也只是对于线性化光照变化敏感, 但条件方差总和(sum of conditional variance, SCV)具有非线性照明变化的不变性<sup>[30]</sup>, 相对于ZNCC而言, 不仅能处理非线性光度变化的情况, 还能避免使用亮度不变性假设; 而对于使用互信息的度量方法, 虽然对照明变化鲁棒, 但计算量比SCV复杂太多。因此使用条件方差总和的方法实现了两者的折中, 能用简单高效的计算达到对照明变化鲁棒的效果。

SCV是一种使用光度的概率分布函数来进行相似性度量的方法<sup>[82]</sup>, 能够动态地使参考图像适应由相机获取的当前图像的照明条件, 考虑当前和期望图像两者中的照明条件, 使用概率期望算子计算出期望图像经过照明变换之后的图像。由于能够动态地调整所需的图像, 即使在残酷的照明变化条件下, 该方法也十分有效, 对于跟踪环境中的任务十分有效。此外, 由于该方法仍然基于像素进行计算, 因此, 相对于互信息而言, 计算效率上有所保障。

相关计算方法如下: 式(30)中的 $P_{II^*}(i, j)$ 表示联合概率密度,  $\hat{\mathbf{I}}(x)$ 表示期望图像 $\mathbf{I}(x)$ 经过照明变换之

后的图像,可以通过SCV计算得到.

$$\begin{cases} \hat{\mathbf{I}}(x) = \mathcal{E}(\mathbf{I}(x) | \mathbf{I}^*(x)) = \sum_i i \frac{P_{II^*}(i, j)}{P_{I^*}(j)}, \\ P_{II^*}(i, j) = \frac{1}{N} \sum_x \phi(\mathbf{I}(x) - i) \phi(\mathbf{I}^*(x) - j), \\ P_{I^*}(j) = \sum_i P_{II^*}(i, j). \end{cases} \quad (30)$$

和以往的伺服任务不同,此时的伺服任务如式(31)所示,即最小化期望图像 $\mathbf{I}(x)$ 与经过照明变化之后图像 $\hat{\mathbf{I}}(x)$ 之间的差异.同时SCV仍然是基于图像强度进行伺服,因此其交互矩阵的形式也相较于互信息而言更为简单.

$$\begin{cases} \text{SCV} = \arg \min_r ((\mathbf{I}(x) - \sum_i i \frac{P_{II^*}(i, j)}{P_{I^*}(j)})), \\ \mathbf{L}_{\text{SCV}} = \frac{\partial \mathbf{I}(r)}{\partial r} = -\nabla \mathbf{I}(r)^T \mathbf{L}_x. \end{cases} \quad (31)$$

## 4 直接视觉伺服技术的控制策略

直接视觉伺服技术关于视觉系统的讨论中,使用新的特征描述方法的目的是找到更合适成本函数,使直接视觉伺服的成本函数具有更好的属性,并最终设计出更有效的控制律.

但选择有效的图像描述方法往往有一定难度.因此,通过将设计控制律视为求解优化问题,将重点置于寻求合适的优化问题的求解方式上,能够减少视觉伺服系统对于选择图像描述方法的依赖,可以实现难点的转换.为此介绍了DVS针对难以求得解析解的问题使用的优化方法,其中包括基本的牛顿法以及相关改进方法,以及使用粒子滤波进行的DVS方法.

### 4.1 非线性优化方法

伺服控制方法主要目的是使得在当前姿势的当前图像中提取的视觉特征达到在期望姿势处获得的期望值,从而正确实现任务.在典型的视觉伺服中,视觉系统往往独立于控制框架,为了能够将视觉信息更好地结合控制器进行设计,Espiau等<sup>[84]</sup>在传感器空间引入了任务函数的概念,并将其用于设计视觉控制器.通过分析大量的考虑视觉信息进行控制器设计的工作,发现几乎所有基于任务函数的视觉伺服,其基本求解步骤可以总结如图4所示的3个过程,在第3节中主要介绍了选择特征与计算交互矩阵这两个环节,针对控制器的设计将在本节中进行介绍<sup>[85]</sup>.

视觉伺服的控制目标在于使得误差向量为零或者任务函数最小化.经典的无标定视觉伺服控制器的设计,主要考虑机器人运动学或机器人动力学<sup>[86]</sup>,包括使用比例积分微分(proportion integration differentiation, PID)控制器<sup>[87]</sup>、使用模型预测控制<sup>[88]</sup>、使用最优控制的框架直接控制力与力矩<sup>[89]</sup>,但基于模型的控制方案往往比较复杂,也出现了学习的手段进行免建

模的伺服工作.但即便如此,对DVS使用以往典型的视觉伺服中的控制方案,会存在收敛率较低的问题.此外,DVS的特点是利用整个图像信息来解决定位任务,其中的控制律不是比较两组几何特征,而是直接分析两个图像之间的相似性度量,这意味着需要一种可靠的方式来进行相似性度量,同时还需设计合理有效的控制方案.

针对DVS的上述特点,出现了一系列方法尝试将视觉伺服过程类比到优化领域.将伺服过程类比于求解非线性最小二乘最小化问题(nonlinear least squares minimization problem, NLS),其中的相似性度量的选择类同于成本函数的构造,控制方案的设计则类同于优化算法的选择.

Collewet等<sup>[90]</sup>尝试类比最优化理论中常见的最小化方法,设计相对应的控制器.当前大多数方法中成本函数 $C(r)$ 多选择式(32)所示的SSD作为相似性度量方法.

$$C(r) = (\mathbf{s}(r) - \mathbf{s}(r^*))^T (\mathbf{s}(r) - \mathbf{s}(r^*)). \quad (32)$$

图6-7展示了基于SSD的成本函数分别在 $(t_x, t_y)$ 和 $(t_x, \theta_y)$ 子空间下的大致形状,可以了解到,实际上,选择好度量方式后其成本函数的一般形式是已知且固定的,但同时成本函数的选择会很大程度上影响收敛域,因此不仅需要选择合适成本函数,也需要结合成本函数选择合适的控制律.

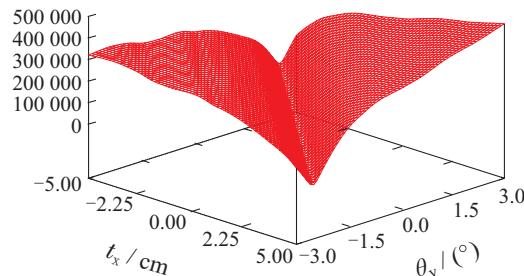


图 7 在 $(t_x, \theta_y)$ 子空间下基于SSD的成本函数<sup>[90]</sup>

Fig. 7 Cost function under a typical subspace in  $(t_x, \theta_y)$

在求解这类NLS的过程中,根据式(33),首先需要选择合适的梯度方向 $\mathbf{d}(r_k)$ 对位姿 $r$ 进行更新.

$$r_{k+1} = r_k \oplus t_k \mathbf{d}(r_k). \quad (33)$$

此外梯度下降的方向需满足式(34)要求:

$$\mathbf{d}(r_k)^T \nabla C(r_k) < 0. \quad (34)$$

为了能够使算法收敛到全局最小值,出现了多种控制方式,在表5<sup>[90-92]</sup>中列出了典型的NLS方法及对应的伺服控制公式即相关特点.

梯度下降法(steepest descent methods, SDM)通过对成本函数进行一阶泰勒展开实现问题的求解,通过朝着梯度相反的方向最小化成本函数.按照始末两种

参考位置, 将SDM按照雅可比矩阵分为常数和变化两种。在机器人控制理论中, SDM对应于雅可比矩阵转

置控制方法(Jacobian transpose control method, JTC), SDM的收敛速度较慢。

表 5 非线性优化方法及其类比的控制器

Table 5 Nonlinear optimization methods and its corresponding controllers

牛顿优化方法	梯度下降方向与控制律设计	特点
SDM <sup>[90]</sup>	$\mathbf{d}(r) = -\nabla \mathbf{C}(r),$ $\mathbf{C}(r) = (\frac{\partial \mathbf{s}}{\partial r})^T (\mathbf{s}(r) - \mathbf{s}(r^*)),$ $\mathbf{v} = -\lambda \mathbf{L}_s^+ (\mathbf{s}(r) - \mathbf{s}(r^*))$	基于一阶泰勒展开进行设计 收敛速度较慢
NMM <sup>[90]</sup>	$\mathbf{d}(r) = -(\nabla^2 \mathbf{C}(r))^{-1} \nabla \mathbf{C}(r),$ $\nabla^2 \mathbf{C}(r) = (\frac{\partial \mathbf{s}}{\partial r})^T (\frac{\partial \mathbf{s}}{\partial r}) + \sum_{i=1}^{i=\dim s} \nabla^2 s_i (s_i(r) - s_i(r^*)),$ $\mathbf{v} = -\lambda \mathbf{L}_s^+ (\mathbf{s}(r) - \mathbf{s}(r^*))$	基于二阶泰勒展开进行设计 对于凹二次函数收敛速度很快 计算量大且部分情况不收敛
GNM <sup>[90]</sup>	$\mathbf{d}(r) = -((\frac{\partial \mathbf{s}}{\partial r})^T (\frac{\partial \mathbf{s}}{\partial r}))^{-1} \nabla \mathbf{C}(r),$ $\mathbf{C}(r) = (\frac{\partial \mathbf{s}}{\partial r})^T (\mathbf{s}(r) - \mathbf{s}(r^*)),$ $\mathbf{v} = -\lambda \mathbf{L}_s^+ (\mathbf{s}(r) - \mathbf{s}(r^*))$	基于NMM方法进行改进 可以实现二次收敛 收敛速度较慢
LMM <sup>[90]</sup>	$\mathbf{d}(r) = -(\mathbf{G} + \mu \text{diag}(\mathbf{G}))^{-1} \nabla \mathbf{C}(r),$ $\mathbf{G} = (\frac{\partial \mathbf{s}}{\partial r})^T (\frac{\partial \mathbf{s}}{\partial r}), \mathbf{H} = \mathbf{L}_s^T \mathbf{L}_s,$ $\mathbf{v} = -\lambda (\mathbf{H} + \mu \text{diag}(\mathbf{H}))^{-1} \mathbf{L}_s^T (\mathbf{s}(r) - \mathbf{s}(r^*))$	相较于GNM收敛速度更快 但不能二次收敛 相较于SDM可线性收敛
ESMM <sup>[91]</sup>	$\mathbf{v} = -2\lambda (\mathbf{L}_s + \mathbf{L}_{s^*})^+ (\mathbf{s}(r) - \mathbf{s}(r^*))$	成本函数的形式已知 使用一阶导数进行二阶近似
MLMM <sup>[92]</sup>	$\mathbf{v}_1 = -v_c \frac{\nabla \mathbf{C}(r_{\text{init}})}{\ \nabla \mathbf{C}(r_{\text{init}})\ },$ $\mathbf{v}_2 = -\lambda (\mathbf{H} + \mu \text{diag}(\mathbf{H}))^{-1} \mathbf{L}_I^T (\mathbf{I}(r) - \mathbf{I}(r^*)),$ $\mathbf{I}(r) = \mathbf{I}(r) + \mathbf{L}_I \Delta r, \mathbf{H} = \mathbf{L}_I^T \mathbf{L}_I$	成本函数的形式已知 结合使用SDM和LMM的思想

牛顿最小化方法(Newton minimization method, NMM)与之对应, 基于成本函数的二阶泰勒级数展开进行求解, 最小化方向则由Hessian矩阵 $\mathbf{H}$ 矩阵的逆来决定。因为NMM的收敛速率是二次的, 如果成本函数为凹二次曲线, NMM将一步到位, 此时的收敛速度比SDM快得多, 但随之也会带来了一定负担的计算量, 此外 $\mathbf{H}$ 矩阵为负定时可能不收敛。

高斯牛顿法(Gauss Newton method, GNM)为解决 $\mathbf{H}$ 矩阵为负定时的不收敛问题, 以损失收敛速度为代价控制 $\mathbf{H}$ 矩阵正定。该方法使用正对称矩阵进行黑森矩阵近似时, 可实现二次收敛, 在机器人控制理论中等价于Jacobian Pseudo-inverse control method(JPC)。

莱文赫格马夸特法(Leven-herg-Marquardt minimization method, LMM)则实现SDM和GNM的光滑变换, 当离最小值较远时, 使用较高的参数 $\mu$ 更易到达谷底, 此时该方法类似于SDM, 但是在这种情况下, 由于谷是窄的, 沿着谷的轴的方向朝向全局最小值的收敛速率是低的, 因此可以选择较低的参数 $\mu$ 来接近全局最小值, 此时该方法类似于GNM, 但也不再确保收敛。在机器人控制理论中, LMM对应于阻尼最小二乘控制法(damped least squares control method, DLSC)。虽然该方法是线性收敛不能实现二次收敛, 但DLSC

技术可以帮助机器人避免奇点问题。

Malis<sup>[91]</sup>提出的有效二阶最小化方法(efficient second order minimization method, ESMM)利用了成本函数的形式是已知的事实, 只使用一阶导数实现对图像误差的有效二阶近似, 避免了成本函数 $r$ 的计算, 也不需要计算 $\nabla^2 \mathbf{C}(r)$ , 和其他与经典的最小化算法相比, 能够很大程度上减少计算量。这种二阶近似能够有效的保证伺服的收敛速度, 对于目标跟踪而言具有天然的优势。

但对于PVS而言, 上述GNM相关的控制律对于以SSD为成本函数的优化问题而言往往容易失败, 其结果要么发散, 要么所得解不适合机器人运动。

Collewet等<sup>[92]</sup>基于此提出了MLMM方法, 该方法利用成本函数的一般形式已知, 首先使用基于梯度的方法设计开环控制器, 并借助范数设计简单的控制律 $\mathbf{v}_1$ 确保实现平滑的3D速度, 随后利用从 $\mathbf{C}(r)$ 的三阶多项式滤波中粗略地检测成本函数谷的底部, 然后利用LM(Levenberg-Marquardt)控制律进行设计, 其中光度特征的交互矩阵使用其近似值 $\mathbf{L}_I$ , 而不是使用成本函数的 $\mathbf{H}$ 矩阵, 由此得到的控制结果 $\mathbf{v}_2$ 。

## 4.2 基于粒子滤波的方法

当前大多DVS使用基于SSD的成本函数进行相似性评估,但从图7可以发现该成本函数是高度非线性的,并且在当前位置可能是非凸的。而实际上,非线性优化方法中,几乎所有伺服控制律都依赖于各种形式的梯度下降来执行优化。当优化的成本函数显示出清晰的凸属性时,这种方法表现非常好,但当凸属性消失时,优化过程可能会失败或卡在局部最小值上或完全发散。因此,为了解决这一局限性,出现了一种优化策略的替代方法,即顺序蒙特卡罗方法也称为粒子滤波方法(particle filtering, PF)。

为了利用粒子滤波器的鲁棒性,Bateux等<sup>[93]</sup>尝试将粒子滤波器集成到视觉伺服控制律中,图8展示了这一基于粒子滤波进行位姿选择的过程。首先从当前位置 $r_0$ 发射出一系列虚拟位置作为粒子,随后结合多视图几何相关知识,利用相机当前位置以及虚拟位置模拟出虚拟图像 $I_k$ ,之后便可以根据成本函数对这些虚拟图像评定,进而估计虚拟位置的优劣也就是估计粒子的质量,并计算每个粒子质量的权重 $w_k$ ,然后选择最佳粒子,并且计算要发送到机器人的相机速度,以便在该最佳粒子即虚拟位置的方向上移动相机。然后对粒子进行重新采样,并重复该过程,直到到达所需的位置。

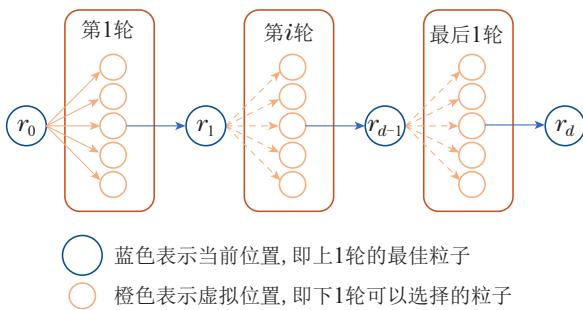


图 8 基于粒子滤波的位姿选择

Fig. 8 Pose selection based on particle filtering

粒子滤波器估计值的实时计算依赖于多视图几何技术和图像传输技术,粒子滤波技术的一般思想是通过一组具有相关权重的随机粒子来表示所需的后验密度函数,并根据这些样本和权重将相机推向最佳状态。因此,能够从虚拟位置预测图像视图非常重要。为了实现这一目标,当前可以设计基于单应性的翘曲函数将当前图像扭曲成虚拟图像。

研究表明该方法基于即使初始化状态在凸区域之外,也能够收敛到全局最小值,主要依赖于成本函数的估计过程,该估计过程通过对搜索空间进行采样并发现经典梯度下降法无法达到的凸区域完成优化。即使对于高度非线性和非凸的成本函数而言,不仅能够出色的完成,还能增加收敛域,这也进一步提高了视觉伺服在成本函数上的使用范围,能够使用的成本函数不限于SSD,也就意味着收敛域增大的可能性在提升,

对于DVS而言,第3节中所述的方法中出现的强大的全局特征,如直方图特征、矩特征、互信息等,其成本函数往往更精细也更难以收敛,均可以用粒子滤波的方式来实现全局最小化提高收敛性<sup>[94]</sup>。

## 5 总结与展望

### 5.1 总结

视觉伺服技术基于视觉的信息量大、适用范围广、非接触性等优势,得到了广泛发展。在机械臂领域的应用与研究一直火热<sup>[21]</sup>,对于移动机器人的视觉伺服技术近年来也开始受到研究者关注<sup>[15-16]</sup>,出现的组合式机器人如移动双臂机械手<sup>[35]</sup>、旋翼飞行机械臂<sup>[20]</sup>的视觉伺服技术同样发展迅速,甚至在卫星跟踪领域也有应用<sup>[18]</sup>。但随之而来的对伺服精度的要求也在逐步提高,典型的视觉伺服控制进行局部特征抽取时,会带来图像处理复杂、伺服控制精度不高、收敛域有限等问题,这也成为了限制视觉伺服发展的一大瓶颈。直接视觉伺服的出现在某种程度上能够解决上述部分问题。

本文从视觉以及控制两个方面综述了面向机器人控制的直接视觉伺服技术。通过介绍典型的视觉伺服框架来引出对直接视觉伺服的讨论。首先从光度这一图像特征的基本描述方法入手,针对光度特征受光照影响大的特点,分析了色彩、深度、梯度作为基本描述方法的优缺点以及相应的交互矩阵形式,随后针对纯光度直接视觉伺服收敛域小等问题,从空间转换角度介绍了基于光度变换的多种特征描述方法,分析介绍了对应方法的优缺点及相关应用,并简要介绍了基本的交互矩阵。之后针对度量特征相似性时所用方法进行了简要介绍。在文章的最后一节总结分析了直接视觉伺服技术中所用控制方法,主要介绍了几类非线性优化方法的特点及相关公式,同时也补充了粒子滤波在直接视觉伺服中的应用。为相关领域的研究者提供了一定参考价值。

### 5.2 展望

近年来,深度学习模型在学习非结构性数据方面的表现越发良好,这得益于深层网络能够学习输入的层次化表示,减少了手工特征提取。在直接视觉伺服上运用深度学习模型,可利用减少中间步骤来有效解决直接视觉伺服中的部分问题,卷积神经网络(convolutional neural networks, CNN)已被证明可以克服直接视觉伺服的部分缺点。

视觉伺服过程中的多数环节都可以用网络直接代替估计。包括使用从目标对象的不同场景捕获的图像及其相应姿势来训练CNN进行位姿估计,再将其用于所得的PBVS方案中<sup>[95]</sup>。但相较于估计姿势而言,视觉伺服中网络还有别的简单的用法如估计交互矩阵<sup>[96-97]</sup>,或者直接利用全局图像信息估计相机速度<sup>[98]</sup>,但是使用全图像进行处理的方式往往因为初期

图像差异较大时容易失败,因此尝试利用基于特征的方法所具有的注意力机制,使用网络自动提取和匹配视觉特征<sup>[99]</sup>,与经典的基于特征的IVS相比,该方法能够直接对特征进行学习,对于未知场景很有帮助,基于网络的视觉伺服还可利用渲染引擎来合成目标图像。此外,使用神经网络可以更好地逼近相机位姿与图像矩的非线性关系<sup>[100]</sup>,上述端到端的方法,不需要机器人的雅可比矩阵,整个过程机器人运动轨迹更光滑,收敛速度更快,鲁棒性也更强。此外,模仿学习与视觉伺服结合也能实现不错效果<sup>[101]</sup>,同时关于图像轨迹的规划问题,也可以通过创建合成数据集的方式,训练网络预测关键帧<sup>[102]</sup>。

## 参考文献:

- [1] SHIRAI Y, INOUE H. Guiding a robot by visual feedback in assembling tasks. *Pattern Recognition*, 1973, 5(2): 99 – 108.
- [2] HUTCHINSON S, HAGER G D, CORKE P I. A tutorial on visual servo control. *IEEE Transactions on Robotics and Automation*, 1996, 12(5): 651 – 670.
- [3] CHAUMETTE F, HUTCHINSON S. Visual servo control. I. Basic approaches. *IEEE Robotics & Automation Magazine*, 2006, 13(4): 82 – 90.
- [4] CHAUMETTE F, HUTCHINSON S. Visual servo control. II. Advanced approaches. *IEEE Robotics & Automation Magazine*, 2007, 14(1): 109 – 118.
- [5] FANG Yongchun. A survey of robot visual servoing. *CAAI Transactions on Intelligent Systems*, 2008, 3(2): 109 – 114.  
(方勇纯. 机器人视觉伺服研究综述. 智能系统学报, 2008, 3(2): 109 – 114.)
- [6] JIA Bingxi, LIU Shan, ZHANG Kaixiang, et al. Survey on robot visual servo control: Vision system and control strategies. *Acta Automatica Sinica*, 2015, 41(5): 861 – 873.  
(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. 自动化学报, 2015, 41(5): 861 – 873.)
- [7] XU De. A tutorial for monocular visual servoing. *Acta Automatica Sinica*, 2018, 44(10): 1729 – 1746.  
(徐德. 单目视觉伺服研究综述. 自动化学报, 2018, 44(10): 1729 – 1746.)
- [8] LI C, LI B, WANG R, et al. A survey on visual servoing for wheeled mobile robots. *International Journal of Intelligent Robotic and Applications*, 2021, 5(2): 203 – 218.
- [9] TAMADAZTE B, PIAT N L F, MARCHAND E. A direct visual servoing scheme for automatic nanopositioning. *IEEE/ASME Transactions on Mechatronics*, 2012, 17(4): 728 – 736.
- [10] SILVEIRA G, MALIS E. Direct visual servoing: Vision-based estimation and control using only nonmetric information. *IEEE Transactions on Robotics*, 2012, 28(4): 974 – 980.
- [11] CHESI G, HASHIMOTO K. *Visual Servoing via Advanced Numerical Methods*. London: Springer London, 2010.
- [12] BATEUX Q. Going further with direct visual servoing. *Computer Vision and Pattern Recognition*, 2018.
- [13] RAMON J L, POMARES J, FELICETTI L. Direct visual servoing and interaction control for a two arms on orbit servicing spacecraft. *Acta Astronautica*, 2022, 192: 368378.
- [14] MEKKI H, LETAIEF M. Path planning for 3D visual servoing: For a wheeled mobile robot. *International Conference on Individual and Collective Behaviors in Robotics (ICBR)*. Sousse, Tunisia: IEEE, 2013: 86 – 91.
- [15] ZHANG Xuebo, FANG Yongchun, LIU Xi. Adaptive visual servo regulation of mobile robots. *Control Theory & Applications*, 2010, 27(9): 1123 – 1130.  
(张雪波, 方勇纯, 刘玺. 移动机器人自适应视觉伺服镇定控制. 控制理论与应用, 2010, 27(9): 1123 – 1130.)
- [16] CAO Yu, LIU Shan. Homography-based platooning control of mobile robots. *Control Theory & Applications*, 2019, 36(9): 1382 – 1390.  
(曹雨, 刘山. 基于单应性矩阵的移动机器人编队跟随控制. 控制理论与应用, 2019, 36(9): 1382 – 1390.)
- [17] CHERUBINI A, CHAUMETTE F, ORIOLO G. An image-based visual servoing scheme for following paths with nonholonomic mobile robots. *Robotics and Vision 10th International Conference on Control, Automation*. Hanoi, Vietnam: IEEE, 2008: 108 – 113.
- [18] ROBIC M, FRAISSE R, MARCHAND E, et al. Vision-based rotational control of an agile observation satellite. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Kyoto, Japan: IEEE, 2022: 2211 – 2218.
- [19] COLLEWET C, CHAUMETTE F. Positioning a camera with respect to planar objects of unknown shape by coupling 2-D visual servoing and 3-D estimations. *IEEE Transactions on Robotics and Automation*, 2002, 18(3): 322 – 333.
- [20] HAFEZ A H A, JAWAHAR C V. Visual servoing by optimization of a 2D/3D hybrid objective function. *Proceedings IEEE International Conference on Robotics and Automation*. Rome, Italy: IEEE, 2007: 1691 – 1696.
- [21] SUN Jingtao, WANG Yaonan, TAN Jianhao, et al. Hybrid visual servoing for rotor aerial manipulation system. *Control Theory & Applications*, 2019, 36(4): 505 – 515.  
(孙敬陶, 王耀南, 谭建豪, 等. 旋翼飞行机械臂系统的混合视觉伺服控制. 控制理论与应用, 2019, 36(4): 505 – 515.)
- [22] LI C, ZHANG X, GAO H, et al. Bridging the gap between visual servoing and visual SLAM: A novel integrated interactive framework. *IEEE Trans Automation Science and Engineering*, 2022, 19(3): 2245 – 2255.
- [23] CONG V D. Visual servoing control of 4DOF palletizing robotic arm for vision based sorting robot system. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 2023, 17(2): 717728.
- [24] GAO Zhendong, SU Jianbo. The estimation of image Jacobian matrix with time-delay compensation for uncalibrated visual servoing. *Control Theory & Applications*, 2009, 26(1): 23 – 27.  
(高振东, 苏剑波. 带有时延补偿的图像雅可比矩阵估计方法. 控制理论与应用, 2009, 26(1): 23 – 27.)
- [25] BATEUX Q, MARCHAND E, LEITNER J, et al. Training deep neural networks for visual servoing. *IEEE International Conference on Robotics and Automation (ICRA)*. Brisbane, QLD: IEEE, 2018: 3307 – 3314.
- [26] MALIS E, CHAUMETTE F, BOUDET S. 2-1/2-D visual servoing. *IEEE Transactions on Robotics and Automation*, 1999, 15(2): 238 – 250.
- [27] LI W, XIONG R. A hybrid visual servo control method for simultaneously controlling a nonholonomic mobile and a manipulator. *Frontiers of Information Technology & Electronic Engineering*, 2021, 22(2): 141154.
- [28] COLLEWET C, MARCHAND E. Photometric visual servoing. *IEEE Transactions on Robotics*, 2011, 27(4): 828 – 834.
- [29] CARON G, MARCHAND E, MOUADDIB E M. Photometric visual servoing for omnidirectional cameras. *Autonomous Robots*, 2013, 35(2/3): 177 – 193.
- [30] DELABARRE B, MARCHAND E. Visual servoing using the sum of conditional variance. *IEEE/RSJ International Conference on Intelligent Robots and Systems*. Vilamoura-Algarve, Portugal: IEEE, 2012: 1689 – 1694.

- [31] SILVEIRA G, MALIS E. Real-time visual tracking under arbitrary illumination changes. *IEEE Conference on Computer Vision and Pattern Recognition*. Minneapolis, MN, USA: IEEE, 2007: 1 – 6.
- [32] COLLEWET C, MARCHAND E. Photometry-based visual servoing using light reflexion models. *IEEE International Conference on Robotics and Automation*. Kobe, Japan: IEEE, 2009: 701 – 706.
- [33] SILVEIRA G, MIRISOLA L, MORIN P. Decoupled intensity-based nonmetric visual servo control. *IEEE Transactions on Control Systems Technology*, 2020, 28(2): 566 – 573.
- [34] HEALEY G. Segmenting images using normalized color. *IEEE Transactions on Systems, Man, and Cybernetics*, 1992, 22(1): 64 – 73.
- [35] YANG Tangwen, GAO Lining, RUAN Qiuqi, et al. Visual servo technology for coordinated manipulation of a mobile dual-arm manipulator system. *Control Theory & Applications*, 2015, 32(1): 69 – 74.  
(杨唐文, 高立宁, 阮秋琦, 等. 移动双臂机械手系统协调操作的视觉伺服技术. *控制理论与应用*, 2015, 32(1): 69 – 74.)
- [36] COLLEWET C, MARCHAND E. Colorimetry-based visual servoing. *IEEE/RSJ International Conference on Intelligent Robots and Systems*. Saint Louis, MO, USA: IEEE, 2009: 5438 – 5443.
- [37] SILVEIRA G, MALIS E. Unified direct visual tracking of rigid and deformable surfaces under generic illumination changes in grayscale and color images. *International Journal of Computer Vision*, 2010, 89(1): 84 – 105.
- [38] SURYA G, SUBBARAO M. Depth from defocus by changing camera aperture: A spatial domain approach. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Seattle, WA, USA: IEEE, 1993: 61 – 67.
- [39] TEULIERE C, MARCHAND E. Direct 3D servoing using dense depth maps. *IEEE/RSJ International Conference on Intelligent Robots and Systems*. Vilamoura-Algarve, Portugal: IEEE, 2012: 1741 – 1746.
- [40] TEULIERE C, MARCHAND E. A dense and direct approach to visual servoing using depth maps. *IEEE Transactions on Robotics*, 2014, 30(5): 1242 – 1249.
- [41] MARCHAND E. Control camera and light source positions using image gradient information. *Proceedings IEEE International Conference on Robotics and Automation*. Rome, Italy: IEEE, 2007: 417 – 422.
- [42] MARCHAND E, COLLEWET C. Using image gradient as a visual feature for visual servoing. *IEEE/RSJ International Conference on Intelligent Robots and Systems*. Taipei, Taiwan: IEEE, 2010: 5687 – 5692.
- [43] BATEUX Q, MARCHAND E. Histograms-based visual servoing. *IEEE Robotics and Automation Letters*, 2017, 2(1): 80 – 87.
- [44] BATEUX Q, MARCHAND E. Direct visual servoing based on multiple intensity histograms. *IEEE International Conference on Robotics and Automation (ICRA)*. Seattle, WA, USA: IEEE, 2015: 6019 – 6024.
- [45] CHAUMETTE F. Image moments: A general and useful set of features for visual servoing. *IEEE Transactions on Robotics*, 2004, 20(4): 713 – 723.
- [46] TAHRI O, TAMTSIA A Y, MEZOUAR Y, et al. Visual servoing based on shifted moments. *IEEE Transactions on Robotics*, 2015, 31(3): 798 – 804.
- [47] BAKTHAVATCHALAM M, CHAUMETTE F, MARCHAND E. Photometric moments: New promising candidates for visual servoing. *IEEE International Conference on Robotics and Automation*. Karlsruhe: IEEE, 2013: 5241 – 5246.
- [48] BAKTHAVATCHALAM M, TAHRI O, CHAUMETTE F. A direct dense visual servoing approach using photometric moments. *IEEE Transactions on Robotics*, 2018, 34(5): 1226 – 1239.
- [49] TAHRI O, CHAUMETTE F. Point-based and region-based image moments for visual servoing of planar objects. *IEEE Transactions on Robotics*, 2005, 21(6): 1116 – 1127.
- [50] LIN Jing, CHEN Huitang, WANG Yuejuan. Visual servoing using image moments. *Journal of Tongji University*, 2000, 28(4): 438 – 442.  
(林靖, 陈辉堂, 王月娟. 基于图像矩的机器人视觉伺服. *同济大学学报*, 2000, 28(4): 438 – 442.)
- [51] TAHRI O, CHAUMETTE F. Application of moment invariants to visual servoing. *IEEE International Conference on Robotics and Automation*. Taipei, Taiwan: IEEE, 2003: 4276 – 4281.
- [52] ABDUL HAFIZ A H, ACHAR S, JAWAHAR C V. Visual servoing based on Gaussian mixture models. *IEEE International Conference on Robotics and Automation*. Pasadena, CA, USA: IEEE, 2008: 3225 – 3230.
- [53] GUERBAS S E, CROMBEZ N, CARON G, et al. Photometric Gaussian mixtures for direct virtual visual servoing of omnidirectional camera. *IEEE CVPR 2021 Workshop on 3D Vision and Robotics*. Seattle (virtual), United States: IEEE, 2021: hal-03227374.
- [54] CROMBEZ N, MOUADDIB E M, CARON G, et al. Visual servoing with photometric Gaussian mixtures as dense features. *IEEE Transactions on Robotics*, 2019, 35(1): 49 – 63.
- [55] MARTINEZ EAR, CARON G, PEGARD C, et al. Photometric-planner for visual path following. *IEEE Sensors Journal*, 2021, 21(10): 11310 – 11317.
- [56] CARON G. Defocus-based direct visual servoing. *IEEE Robotics and Automation Letters*, 2021, 6(2): 4056 – 4063.
- [57] CUI L, MARCHAND E, HALIYO S, et al. Hybrid automatic visual servoing scheme using defocus information for 6-DoF micropositioning. *IEEE International Conference on Robotics and Automation (ICRA)*. Washington, USA: IEEE, 2015: 6025 – 6030.
- [58] DEGUCHI K. A direct interpretation of dynamic images with camera and object motions for vision guided robot control. *International Journal of Computer Vision*, 2000, 37: 7 – 20.
- [59] NAYAR S K, NENE S A, MURASE H. Subspace methods for robot vision. *IEEE Transactions on Robotics and Automation*, 1996, 12(5): 750 – 758.
- [60] MARCHAND E. Subspace-based direct visual servoing. *IEEE Robotics and Automation Letters*, 2019, 4(3): 2699 – 2706.
- [61] MARTURI N, TAMADAZTE B, DEMBELE S, et al. Visual servoing schemes for automatic nanopositioning under scanning electron microscope. *IEEE International Conference on Robotics and Automation (ICRA)*. HongKong, China: IEEE, 2014: 981 – 986.
- [62] GUELPA V, LAURENT G J, TAMADAZTE B, et al. Single frequency-based visual servoing for microrobotics applications. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Daejeon, Korea: IEEE, 2016: 4274 – 4279.
- [63] AHMED N, NATARAJAN T, RAO K R. Discrete cosine transform. *IEEE Transactions on Computers*, 1974, 23(1): 90 – 93.
- [64] OURAK M, TAMADAZTE B, LEHMANN O, et al. Direct visual servoing using wavelet coefficients. *IEEE/ASME Transactions on Mechatronics*, 2019, 24(3): 1129 – 1140.
- [65] OURAK M, TAMADAZTE B, LEHMANN O, et al. Wavelets-based 6 DOF visual servoing. *IEEE International Conference on Robotics and Automation (ICRA)*. Stockholm, Sweden: IEEE, 2016: 3414 – 3419.
- [66] DUFLOT L A, KRUPA A, TAMADAZTE B, et al. Towards ultrasound-based visual servoing using shearlet coefficients. *IEEE International Conference on Robotics and Automation (ICRA)*. Stockholm, Sweden: IEEE, 2016: 3420 – 3425.
- [67] DUFLOT L A, REISENHOFER R, TAMADAZTE B, et al. Wavelet and shearlet-based image representations for visual servoing. *The International Journal of Robotics Research*, 2019, 38(4): 422 – 450.
- [68] ZHAO J, SONG B, XI N, et al. Compressive feedback based non-vector space control. *American Control Conference (ACC)*. Montreal, Canada: IEEE, 2012: 4090 – 4095.

- [69] SONG B, ZHAO J, XI N, et al. Compressive feedback-based motion control for nanomanipulation-theory and applications. *IEEE Transactions on Robotics*, 2014, 30(1): 103 – 114.
- [70] LI C, WANG S, WANG S, et al. Robot motion control with compressive feedback. *IEEE International Conference on Robotics and Automation (ICRA)*. Xi'an, China: IEEE, 2021: 12925 – 12930.
- [71] FELTON S, BRAULT P, FROMONT E, et al. Visual servoing in autoencoder latent space. *IEEE Robotics and Automation Letters*, 2022, 7(2): 3234 – 3241.
- [72] HAGER G D, DEWAN M, STEWART C V. Multiple kernel tracking with SSD. *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. Washington, DC, USA: IEEE, 2004: 790 – 797.
- [73] KALLEM V, DEWAN M, SWENSEN J P, et al. Kernel-based visual servoing. *IEEE/RSJ International Conference on Intelligent Robots and Systems*. San Diego, CA, USA: IEEE, 2007: 1975 – 1980.
- [74] CHEN Y, MENG Q H, LIU L. Direct visual servoing based on discrete orthogonal moments. *ArXiv Preprint*, 2023, abs/2304.14012.
- [75] MARCHAND E. Direct visual servoing in the frequency dDomain. *IEEE Robotics and Automation Letters*, 2020, 5(2): 620 – 627.
- [76] IRANI M, ANANDAN P. Robust multi-sensor image alignment. *The 6th International Conference on Computer Vision*. Bombay, India: IEEE, 1998: 959 – 966.
- [77] DAME A, MARCHAND E. Entropy-based visual servoing. *IEEE International Conference on Robotics and Automation*. Kobe, Japan: IEEE, 2009: 707 – 713.
- [78] DAME A, MARCHAND E. Mutual information-based visual servoing. *IEEE Transactions on Robotics*, 2011, 27(5): 958 – 969.
- [79] DAME A, MARCHAND E. Improving mutual information-based visual servoing. *IEEE International Conference on Robotics and Automation*. Anchorage, Alaska, USA: IEEE, 2010: 5531 – 5536.
- [80] BARTOLI A, HAGER G D, BOUTHEMY P, et al. *A Unified Direct Approach for Visual Servoing and Visual Tracking Using Mutual Information*. Université Rennes 1, 2010.
- [81] RAJ S, GIORDANO P R, CHAUMETTE F. Appearance-based indoor navigation by IBVS using mutual information. *The 14th International Conference on Control, Automation, Robotics and Vision (ICARCV)*. Phuket, Thailand: IEEE, 2016: 1 – 6.
- [82] RICHA R, SZNITMAN R, TAYLOR R, et al. Visual tracking using the sum of conditional variance. *IEEE/RSJ International Conference on Intelligent Robots and Systems*. San Francisco, California, USA: IEEE, 2011: 2953 – 2958.
- [83] LI X, ZHAO H, DING H. Kullback-Leibler divergence based visual servoing. *IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*. Delft, Netherlands: IEEE, 2021: 720 – 726.
- [84] ESPIAU B, CHAUMETTE F, RIVES P. A new approach to visual servoing in robotics. *IEEE Transactions on Robotics and Automation*, 1992, 8(3): 313 – 326.
- [85] ZHANG Guoliang, XIE Zongwu, WANG Jie, et al. Study of visual servo-performance of robots based on task function. *Control Theory & Applications*, 2009, 26(11): 1197 – 1203.  
(张国亮, 谢宗武, 王捷, 等. 基于任务函数方法的机器人视觉伺服特性研究. 控制理论与应用, 2009, 26(11): 1197 – 1203.)
- [86] TAO Bo, GONG Zeyu, DING Han. Survey on uncalibrated robot visual servoing control. *Chinese Journal of Theoretical and Applied Mechanics*, 2016, 48(4): 767 – 783.  
(陶波, 龚泽宇, 丁汉. 机器人无标定视觉伺服控制研究进展. 力学学报, 2016, 48(4): 767 – 783.)
- [87] LIN Jing, CHEN Huitang, WANG Yuejuan, et al. Research on robotic visual servoing system. *Control Theory & Applications*, 2000, 17(4): 476 – 482.  
(林靖, 陈辉堂, 王月娟, 等. 机器人视觉伺服系统的研究. 控制理论与应用, 2000, 17(4): 476 – 482.)
- [88] XING Kexin, NI Weiqi, HE Defeng. Robust predictive control of visual servoing systems of mobile robots with load uncertainty. *Control Theory & Applications*, 2022, 39(2): 327 – 335.  
(邢科新, 倪伟琦, 何德峰. 荷载不确定移动机器人视觉伺服系统鲁棒预测控制. 控制理论与应用, 2022, 39(2): 327 – 335.)
- [89] BELMONTE A, RAMON J, POMARES J, et al. Optimal image-based Guidance of mobile manipulators using direct visual servoing. *Electronics*, 2019, 8(4): 374.
- [90] COLLEWET C, MARCHAND E, CHAUMETTE F. Asservissement visuel basé sur des informations photométriques. *16e Congrès Francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle*. Amiens, France: Rennes 1, 2008: 435 – 443.
- [91] MALIS E. Improving vision-based control using efficient second-order minimization techniques. *Proceedings IEEE International Conference on Robotics and Automation*. New Orleans, LA, USA: IEEE, 2004: 1843 – 1848.
- [92] COLLEWET C, MARCHAND E, CHAUMETTE F. Visual servoing set free from image processing. *IEEE International Conference on Robotics and Automation*. Pasadena, CA, USA: IEEE, 2008: 81 – 86.
- [93] BATEUX Q, MARCHAND E. Particle filter-based direct visual servoing. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Daejeon, Korea: IEEE, 2016: 4180 – 4186.
- [94] BATEUX Q, MARCHAND E. Histograms-based visual servoing. *IEEE Robotics and Automation Letters*, 2017, 2(1): 80 – 87.
- [95] BATEUX Q, MARCHAND E, LEITNER J, et al. Visual servoing from deep neural networks. *ArXiv Preprint*, 2017, arXiv: 1705.08940.
- [96] BATEUX Q, MARCHAND E, LEITNER J, et al. Training deep neural networks for visual servoing. *IEEE International Conference on Robotics and Automation (ICRA)*. Brisbane, QLD: IEEE, 2018: 3307 – 3314.
- [97] PRZYSTUPA M, DEHGHN M, JAGERSAND M, et al. Analyzing neural jacobian methods in applications of visual servoing and kinematic control. *ArXiv Preprint*, 2021, arXiv: 2106.06083.
- [98] ABDULHAFIZ I, NAZARI A A, ABBASI-HASHEMI T, et al. Deep direct visual servoing of tendon-driven continuum robots. *IEEE 18th International Conference on Automation Science and Engineering (CASE)*. Mexico City, Mexico: IEEE, 2022: 1977 – 1984.
- [99] ADRIAN N, DO V T, PHAM Q C. DFBVS: Deep feature-based visual servo. *ArXiv Preprint*, 2022, arXiv: 2201.08046.
- [100] LI Youxin, MAO Zongyuan, TIAN Lianfang. Visual servoing of 4D-OF using image moments and neural network. *Control Theory & Applications*, 2009, 26(10): 1162 – 1166.  
(李优新, 毛宗源, 田联房. 基于图像矩与神经网络的机器人四自由度视觉伺服. 控制理论与应用, 2009, 26(10): 1162 – 1166.)
- [101] PAOLILLO A, SAVERIANO M. Learning stable dynamical systems for visual servoing. *ArXiv Preprint*, 2023, arXiv: 2204.05681.
- [102] CROMBEZ N, BUISSON J, YAN Z, et al. Subsequent keyframe generation for visual servoing. *IEEE International Conference on Robotics and Automation (ICRA)*. Xi'an, China: IEEE, 2021: 14439 – 14445.

### 作者简介:

**黄梦洁** 硕士研究生, 目前研究方向为机器视觉、机器人视觉伺服控制, E-mail: mengjie.huang@foxmail.com;

**叶磊** 副研究员, 博士, 目前研究方向为自动驾驶视觉感知、路径规划、智能控制, E-mail: yelei@nudt.edu.cn;

**易凡骁** 硕士研究生, 目前研究方向为无人驾驶车辆路径规划技术, E-mail: yi\_fanxiao@foxmail.com;

**王千** 硕士研究生, 目前研究方向为无人驾驶车辆目标检测技术, E-mail: wang.qian.slzx@foxmail.com.