

本刊专稿

具身机器人遥操作综述

倪自强, 谢绍玄, 柳学成, 杨翔, 刘优, 姚国才*

摘要 具身机器人遥操作在机器人完全实现类人的自主化之前, 还将长期存在。首先, 分类讨论了遥操作的人机交互方式, 包括外骨骼、虚拟/混合现实、动作捕捉/惯性传感, 以及视觉等接口方案。其次, 分析了遥操作控制模式的演变路径, 涵盖直接控制、共享控制、基于模仿学习的控制及生成式策略的引入, 阐明各自的原理与应用场景。再者, 介绍了典型遥操作系统的体系架构与技术实现, 涵盖感知与映射、控制算法与通信模块, 并探讨视觉、力觉、触觉等多模态反馈在提升沉浸体验中的作用。进一步, 梳理了遥操作系统性能评估的方法与最新基准, 指出开源软硬件平台和数据资源在推动研究中的重要意义。最后, 总结了当前面临的挑战与未来发展方向, 包括智能化水平提升、成本与使用门槛的降低, 以及标准体系的建立等。

关键词 具身机器人; 遥操作; 人机交互; 控制模式

具身机器人指其物理形态(形体、材料、传感与致动配置)与所处环境的相互作用对其感知、控制与“智能”行为具有构成性作用的机器人系统, 其认知与控制并非可脱离身体而独立存在, 而是由身体-环境-控制三者耦合共同涌现^[1-2]。具身机器人遥操作是指人类操作员通过特定接口远程控制机器人完成任务, 使机器人如同操作者的“化身”, 在本地/异地执行复杂操作^[3]。尽管人工智能与机器人自主技术发展迅速, 完全自主的机器人在开放环境中仍难以实现与人类同等的灵巧行为^[4]。因此, 遥操作作为连接人类智能与机器人体能的桥梁, 在实现通用人工智能(artificial general intelligence, AGI)之前仍将发挥重要作用, 尤其在建筑施工、化工厂、核污染区、太空等人类难以直接作业的高危或不可达场景下具有不可替代的价值^[5]。通过遥操作, 机器人可替代人类进入这些高风险环境执行检查、维护等任务, 避免人员直接暴

露于危险之中。与此同时, 在具身智能发展的早期阶段, 遥操作还常被用于具身数据的采集, 为后续机器人学习与自主能力提升提供基础。但需要注意: 遥操作示教数据往往带有明显的操作者行为偏好(如速度/力度习惯、注视与视角选择、路径规划风格、纠错方式等), 且在不同操作者、不同疲劳水平与不同设备条件下, 数据分布会出现漂移, 导致数据质量稳定性下降。为降低偏好与漂移对训练的影响, 采集阶段建议引入: (1) 操作者分层(新手/熟练/专家)与跨人群混采; (2) 同任务多次重复与跨时段采集; (3) 关键质量指标(成功率、轨迹平滑度、接触力范围、丢帧率、时延)在线/准在线监控; (4) 对“异常演示”(过度抖动、过度停顿、频繁失败)进行自动剔除或降权; (5) 在数据集中显式记录操作者与设备元信息(接口类型、延迟、视角配置、是否有力反馈)以支持后续分组训练与去偏。此外, 遥操作还克服了地域限制, 例如, 支持远程医疗手术和护理服务, 使专家无须亲临现场即可提供操作与协助。特别是类人型的具身机器人由于

北京智源人工智能研究院, 北京 100089

收稿日期: 2025-09-23; 修回日期: 2025-12-31

作者简介: 倪自强, 高级工程师, 研究方向为具身机器人, 电子信箱: zqni@baai.ac.cn; 姚国才(通信作者), 高级工程师, 研究方向为具身智能, 电子信箱: gcyao1@baai.ac.cn

引用格式: 倪自强, 谢绍玄, 柳学成, 等. 具身机器人遥操作综述[J]. 科技导报, 2026, 44(13): 28-39; doi:[10.3981/j.issn.1000-7857.2025.09.00103](https://doi.org/10.3981/j.issn.1000-7857.2025.09.00103)

形态与人类相似,更适合在人类环境中使用工具并与人互动^[3,6];在远程护理等需要与人直接接触的应用中,这类机器人也展现出独特优势^[4]。

1 遥操作交互接口分类与比较

在具身机器人遥操作系统中,操作端的人机交互图 1 为典型接口的遥操作系统示例。

表 1 遥操作交互接口的特点比较

接口类型	优点	缺点	典型应用/示例
外骨骼系统	控制直观,一对一映射精确;可提供双边力反馈,沉浸感强	装置笨重昂贵,穿戴复杂;长时间全身操作易疲劳,通常仅用于上肢或特定关节	精细力控场景,如远程手术、重物搬运; iCub3 Avatar ^[8] 外骨骼系统; HOMIE ^[9] 外骨骼系统
VR(虚拟现实)/MR(混合现实)接口	设备廉价易部署;通用性强,可适配不同机器人;操作者以第一视角控制	无力觉反馈,难判断操作力度;视觉跟踪易受环境影响;仅单向控制,安全性稍弱	通用遥操作和远程呈现; Open-TeleVision ^[10] , Bunny-VisionPro ^[11] , Mobile-TeleVision ^[12] , OpenWBT ^[13] , CLONE ^[10]
动捕/IMU(惯性测量单元)接口	获取全身高自由度动作;无须施加机械力,较舒适	光学动捕昂贵且不便携;IMU需复杂校准且易受干扰;无力反馈,长时操作疲劳	需要高保真动作映射的场景(舞蹈、体育动作模仿); TWIST ^[14] 动捕全身控制
视觉跟踪接口	不需要穿戴设备,用户友好;成本极低,易于大规模应用	精度受限(遮挡、光照影响大);仅能提供末端目标控制,无细节姿态和力反馈	远程示教数据采集、人机共融控制; OmniH2O ^[15] 使用单目相机+语音进行基本操控
等比例机械臂主控	所见即所得的直觉控制;天然避免超出机器人能力的指令;可实现双边力反馈,不需要复杂算法	装置需按机器人比例定制,通用性差;模型缩小导致分辨率和力反馈精度受限	高同步性双手操作、装配任务; CHILD ^[16] 系统提供可重构的便携式主从机器人



图 1 不同交互接口的遥操作系统

1.1 外骨骼式接口

外骨骼设备通过机械结构绑定在人身上,检测人体关节运动并同步驱动远程机器人相应关节。外骨骼方式的优点是映射关系直接、控制直观,能实现高

沉浸感和双边力反馈。这使操作者仿佛亲身接触远程环境。但外骨骼系统往往体积庞大、造价高昂,需要复杂的穿戴调试,操作者长时间全身运动也容易疲劳。因此,目前外骨骼多用于上半身控制或特定关节控制,并常结合其他装置辅助下肢移动等。典型案例是意大利技术研究院(Istituto italiano di tecnologia, Iit)开发的 iCub3 Avatar^[8]系统:操作者穿戴全身外骨骼和 VR 设备,可以远程操控人形机器人行走和操作,并通过外骨骼获得视觉、听觉、触觉、重量等全面反馈,实现“完全沉浸”的远程化身。外骨骼接口适用于需要精细力控的遥操作场景,如医疗手术机器人、重物搬运等。还有低成本外骨骼的探索,例如,OpenRobotLab 的 HOMIE^[9]系统使用价廉的等构上肢外骨骼和手套,与强化学习控制下肢相结合,实现全身控制的同时将硬件成本降至仅 500 美元。

1.2 虚拟现实与混合现实接口

利用 VR 或 MR 设备进行遥操作是一种灵活且日益普及的方案。操作者佩戴 VR 头盔获得机器人的第一视角视觉,并使用手持控制器、数据手套等输入设备来发出指令。VR/MR 接口的突出优点是设备相对廉价、部署方便,操作者的运动捕获可以只依赖消费级的 VR 设备和摄像头。其高度通用性意味着同一套 VR 遥操作系统可以适配不同的机器人,只需要调整末端映射算法,不需要像外骨骼那样针对特定机型定制。然而,VR/MR 通常只能提供单向控制,操作者将指令发送给机器人,却无法像外骨骼那样接收机器人物理反馈^[3,8,17]。缺乏力觉反馈会降低操作的直观性和安全性。例如,当机器人手爪抓取到物体时,操作者难以凭感觉判断抓取力度是否合适。此外,VR/MR 依赖的视觉定位可能受遮挡、环境光等影响导致精度下降^[18]。尽管如此,得益于成本和易用性优势,VR/MR 接口在遥操作研究中应用广泛。例如,上海 AI 实验室的 TeleOpBench 基准就将典型的 VR 设备操作作为 4 类代表性遥操作模式之一^[19]。

1.3 动作捕捉与惯性传感接口

动作捕捉(motioncapture, MoCap)和惯性传感单元(inertial measurement unit, IMU)技术通过捕获人体关键点运动来驱动机器人^[20-21]。常见做法是在操作者身上穿戴带反光标记点的动捕服,由红外摄像机阵列实时获取人体骨架关节角度;或佩戴一套 IMU 传感

器,利用惯性测量实现全身姿态捕获。这类接口与外骨骼相似,都能获取操作者全身多自由度动作,用于同步控制人形机器人全身关节。不同在于 MoCap/IMU 不对人体施加机械力,因而设备更轻便,对人的拘束更小。然而其缺点是需要专门的传感器设备,专业光学动捕系统价格昂贵且安装烦琐,通常固定在实验室中,便携性不够。IMU 方案虽然没有空间限制,但同样需要多个传感器标定,易受磁场干扰影响精度。总体而言,动捕/IMU 接口在保证控制直观性的同时简化了设备,但依旧缺乏力反馈和长期操作的舒适性。一些研究通过结合额外的步态控制器来辅助稳定,例如,在操作者腿部动作基础上由算法保证机器人行走平衡。动捕/IMU 适用于需要高保真度动作映射的场景,如让机器人模仿人类的复杂舞蹈动作或体育动作等^[14]。伊利诺伊大学厄巴纳-香槟分校(University of Illinois Urbana-Champaign, UIUC)的 CHILD^[16]等工作倾向于用更简便的传感手段(如腰部 9 轴 IMU 配合上半身机械臂控制)实现全身遥操作,以降低对昂贵动捕系统的依赖。

1.4 视觉跟踪接口

视觉跟踪是指通过普通摄像头来捕捉操作者的动作姿态,从而控制远程机器人。典型方案包括利用 RGB-D 相机(深度摄像头)或单目摄像头结合计算机视觉算法估计人体骨架。例如,使用 OpenPose^[22]、MediaPipe 等^[23]人体姿态估计算法,从摄像头画面中实时获得操作者手臂的关键点位置,再通过逆运动学映射到机器人的手臂关节。视觉接口的优点是不需要穿戴设备,用户友好度高,只要有摄像头即可进行远程操作演示。这使得系统成本极低,且可以方便地对不特定用户开放。然而其局限也很明显:受限於视觉识别精度,人体姿态估计在有遮挡、光线不佳时可能产生较大误差^[19],导致机器人动作不准确。此外,纯视觉跟踪通常只能提供机器人末端目标位姿控制,缺乏对姿态细节和用力程度的感知,且同样没有触觉/力觉反馈^[15,21]。因此,仅靠视觉接口难以完成精细操作或需要闭环力控的任务。总之,视觉接口更适合作为一种远程示教手段,用于快速收集人类演示数据以训练机器人技能。例如,OmniH2O^[15]框架允许操作者仅用单目相机和语音命令对机器人进行基本操控,再由学习策略接管执行复杂动作。这些都体现了视觉接

口在数据收集和人机协作控制中的价值。

1.5 等比例机械臂主控

近年来,一种新兴的遥操作接口是使用等比例或者缩小的机械“主机器人”(leader)来控制远程“从机器人”(follower)。操作者可以直接操作一个等比例或者小型机器人,各个关节的运动通过一对一映射发送给远程的大型机器人,实现关节空间的直接控制。例如,CHILD^[16]系统提供了一个可背负在操作员胸前的小型人形机器人骨架,包含与真实机器人对应的关节和连杆。操作员用双手移动该迷你机器人四肢,各关节角度经比例缩放后发送给远程的 UnitreeG1 全尺寸人形机器人,使其全身动作与操作者同步。这种方式的优点在于控制直观,操作者相当于在“摆弄”一个缩小版机器人,自然地感知到关节极限、奇异姿态等约束,避免发出超出机器人能力范围的指令。其缺点是 Leader 机械结构需要针对特定机器人构型制作,泛用性略差;并且分辨率和力反馈精度受制于缩小模型的性能。不过,已有研究证明这种方法在双腕精细操作中效果良好,并正逐步扩展到全身控制领域。CHILD 是一款面向全身人形机器人的直接关节映射遥操作系统,采用模块化可重构的设计,可通过不同关节安装配置适配多种机器人。总的来看,等比例机械主控为遥操作提供了一种“所见即所得”的直接映射方式,在双手操作、精细装配等需要高同步性的任务中显示出巨大潜力。

从“数据用于训练”视角,不同遥操作接口采集的数据分辨率、噪声结构与可扩展性存在显著差异,因

而更适合不同训练范式。(1) 外骨骼/等比例主从: 关节空间映射明确、可引入力/触觉反馈,易获得高精度动作与接触信息,更适合精细操作的行为克隆 (behavior cloning, BC)/轨迹级模仿学习与需要接触约束的策略学习;(2) VR/MR: 第一视角数据易规模化、任务覆盖面广,适合构建多任务数据集,更适合训练通用视觉语言动作模型(vision-language-action, VLA)/多任务策略与“语言/目标条件”控制;(3) MoCap/IMU: 更擅长全身姿态捕获,适合“全身协同/移动操作”场景,更适合训练全身运动策略(loco-manipulation)或分层控制中的高层意图模型;(4) 纯视觉跟踪: 部署成本最低、可众包,但噪声与遮挡问题更突出,更适合“弱标注/大规模数据驱动”的训练路线,并建议配合后处理(平滑、对齐、异常检测)与策略鲁棒化训练。

2 控制模式的演进

随着具身机器人遥操作技术的发展,控制模式朝着提高机器人自主性并降低人类操作负担的方向不断演进。早期的遥操作主要采用完全依赖人工操控的直接控制,随后发展出人机协作的共享控制模式。此后,通过引入人类示范数据又催生了模仿学习方法,最终形成了利用生成式模型引导机器人行为的高级控制方式。这些控制模式之间具有明显的进化关系,从全人工操控逐步走向高度自主。表 2 分别介绍这些控制模式及其主要特点。

表 2 遥操作控制模式的特点比较

控制模式	代表	人/机分工	优势	局限性
直接控制	DRC, da Vinci ^[24-25]	人直接控制全部机器人动作;机器人无自主性	操作直观,人类灵活应变;适用于未知环境复杂任务	操作员负担重,需要同时管理众多自由度;效率低,易出错,不适合长时间或高延迟场景
共享控制	Centauro, HOMIE ^[26-28]	人决策高层意图+机器人自主优化细节(或监督介入)	降低人负担,结合机器人自主提高成功率和安全性;实现人机协作效应	算法复杂,需要准确预测人意图;人机权重融合不当可能引入干扰;需要高度信任和协调
模仿学习控制	TRILL, Diffusion Policy, H2O/OmniH2O ^[15,29-31]	人示范收集数据,机器人离线学习策略,再自主执行类似任务	利用人工演示快速获取复杂技能;减少在线人工干预,实现一定自主性	训练策略需大量多样示范;泛化能力有限,仍需要人工监控纠偏;无法实时适应新情况
生成式策略控制	Mobile-TeleVision, OmniH2O ^[15,32]	人提供高层目标, AI大模型生成动作方案,机器人自主执行细节	AI规划能力强,可理解抽象指令;有望大幅降低人工操作细节负担	尚处早期,可靠性和安全性未充分验证;对模型和计算依赖高,一旦出错难以纠正

2.1 直接控制

直接控制是由人类操作员全权操纵机器人的运动。操作员的每个动作指令都实时发送给机器人执行,机器人本身不具有自主决策能力。直接控制的优点在于人具有充分的掌控力和灵活性,可以根据实时感知灵活调整机器人行为,适用于未知环境下需要人工机动应变的任务。在诸如复杂装配、救灾场景等无法预先编程的情况下,完全的人控可避免不确定性带来的风险。许多早期遥操作系统(如核工业机械臂遥控、传统手术机器人)均采用这种模式^[33]。

但直接控制的局限性也非常明显。首先,操作者必须同时关注机器人的所有自由度,这对人的认知和反应速度是极大考验。尤其对于人形机器人,操作员需要同时控制双手、双脚的位置和平衡等,纯人工控制常导致机器人动作笨拙、效率低下,甚至频繁失败。研究显示,仅依赖人工经验直接遥控复杂任务,往往需要多次尝试才能成功。其次,人在回路中会引入人为失误和反应延迟,当通信有时延或带宽受限时,纯人控容易因信息滞后导致操作不稳。再次,长时间全手动操作对人的体力和精神是双重负担,稍有不慎就可能出現误操作事故。美国国防部高级研究计划局机器人挑战赛(DARPA robotics challenge, DRC)中各团队的实践也表明,即使专家远程控制,人形机器人完成复杂任务也往往步履维艰,需要借助半自主功能提高效率。总的来说,直接控制模式虽然直观,但在复杂任务中效率和可靠性受限,因而推动了共享控制等改进模式的兴起^[34-35]。

2.2 共享控制

共享控制(shared control)通过引入一定程度的机器人自主性,缓解了直接控制下操作者负担过重的问题。在共享控制框架中,人和机器人成为协作伙伴,共同决定机器人的动作:操作员提供高层次意图或初始指令,机器人控制系统根据自身传感和模型对人类指令进行优化或适当修正,从而更高效地完成任务。典型案例是在双臂操作中引入“在环助手”——当操作员用上肢遥控抓取笨重物体时,机器人自动平衡下肢站姿,并在末端执行器逼近目标时提供对准和稳定辅助。Rakita等^[36]的研究中,操作者控制人形机器人上肢抓取物体,系统根据末端偏差实时校正臂的轨迹,降低对精准操作者输入的依赖,提高成功率。另

一个典型案例是 Rahal等^[37]开发的遥操作切割系统,该系统在操作者大致沿切线移动时,通过引入非完整性约束(nonholonomic constraints),机器人自动沿工件表面进行修正,确保切割稳定,这属于共享控制在特定任务中的应用。

共享控制的一般实现方式包括意图预测与指令融合。系统首先根据操作员的输入序列推测其意图(如目标位置),并据此生成一条辅助运动轨迹。随后,这条辅助轨迹与操作员的原始输入按照一定的权重进行融合,形成最终的控制指令。权重的大小可以根据预测的置信度动态调整,例如,当机器人逐渐接近推测目标时,系统会逐步增大自主控制的比重,从而实现自然的接管过程。这种“人机折中”策略已成功用于机械臂的遥操作导航和视觉伺服等场景。另有一种共享控制形式是监督/守护模式,在机器人具备较高自主性的情况下,由人做监控者,一旦自动行为出现偏差再介入修正。例如,在商场机器人试验中,多个服务机器人自行对话、规划路径,远程人类仅通过界面监视,当机器人遇到未建模障碍或复杂询问时才接管操作。反之亦然,有研究在操作无人车时引入安全守护进程:正常情况完全听从人类遥控,若探测到碰撞危险则守护程序会优先避障,即使覆盖人工指令也要确保系统安全。类似思路可用于人形机器人防止其跌倒或运动超出稳定范围。

在具身机器人遥操作领域,出现了半自主遥操作的实践,即将机器人某些子系统交由智能算法控制,人则专注其余部分,从而实现优势互补。HOMIE^[9]系统就是典型案例:操作员通过外骨骼舱控制机器人上肢和手部,而机器人下肢的行走、蹲起由预先训练的强化学习策略自动执行。操作者只需要踩下踏板给出移动意图,机器人便自主调整步态保持平衡和高度,使上半身操作更平稳。通过这种人机分工,HOMIE在各种操作任务中效率提升近1倍。又如 Mobile-TeleVision^[12]方法提出将人形机器人上肢操作与下肢移动的控制解耦:上半身动作由预测模型产生辅助参考,下半身由独立的闭环控制,避免两者相互干扰。实验证明该方法在保证步态稳定的同时显著提高了上肢操作精度。由此可见,共享控制通过多种途径提升了遥操作的性能和安全性,使人机协作达到“1+1>2”的效果。

2.3 模仿学习控制

模仿学习控制将遥操作和机器人自主学习相结合,旨在让机器人学会人类示范的技能,从而逐步减少对人直接操控的依赖。其典型流程是:由人通过遥操作示范完成一系列任务,记录下机器人传感器和执行器状态轨迹,形成示范数据集;然后采用机器学习算法(如行为克隆、逆强化学习等)训练机器人的控制策略,使其能够在没有人控制时重现类似行为。由于人类在回路中提供了高质量的演示,这种方法能够极大降低机器人自主学习的样本需求^[38-39]。大量研究表明,人类遥操作演示数据是训练机器人实现复杂技能的宝贵资源。例如,在单臂和双臂机器人抓取、操纵任务上,遥操作产生的大规模演示库使深度强化学习或模仿策略得以快速收敛,成功率明显提升^[20-21]。

在人形机器人场景,模仿学习控制尤为重要。一方面,全身遥操作数据直观体现了完成任务所需的人形运动协调模式,包含大量隐含的力学和策略信息;另一方面,人形机器人自主探索成本高、风险大,通过离线学习人类示范可以避免机器人反复试错造成损坏。近期的 TRILL 方法让人远程操控人形机器人完成搬箱、爬梯等杂技般的整身任务,收集示范后训练出深度神经策略,使机器人在无人操控时也能完成类似动作^[29]。又如 Chi 等^[31]提出的扩散策略(diffusion policy)将人类遥操作数据用于生成模型训练,学习到人形机器人操作的隐式分布。通过在真实机器人上进行上千次遥操作滚动(rollout),该方法积累了丰富的状态-动作序列,再利用扩散模型进行有噪声数据的学习和规划,最终使机器人即便只在单一场景获取数据,也能泛化到多个新场景自主完成任务^[30]。值得注意的是,模仿学习控制并不意味着完全排除人,很多方案采用人类指导下的训练(如 DAgger 算法,让人在线修正机器动作^[40])来进一步提升策略性能。这体现了遥操作与自主学习的融合趋势:人在初期提供示范,随后逐步“教会”机器人自主执行,从而将人从持续低层控制中解放出来。

2.4 生成式策略控制

随着人工智能生成模型(generative model)的兴起,将其引入遥操作控制成为新趋势。一种思路是利用生成模型强大的预测与规划能力,辅助或替代人工制定机器人动作。例如,前述 Mobile-TeleVision 方

法中训练的条件变分自编码器(conditional variational autoencoders, CVAE)模型可根据当前状态生成未来一段时间的上肢运动先验,从而引导下肢控制策略协调动作。这种生成式运动先验使机器人能够在保持稳定步态的同时完成复杂手部操作,相当于为机器人提供了“预见性”的运动规划能力^[12]。另一种更高层的生成式控制是结合大模型(如大型语言模型(large language model, LLM))来理解和执行人类意图。Omni H2O 系统展示了这样的雏形:它允许操作员用自然语言(语音)对机器人下达指令,并借助类似 GPT-4 的大模型将高层指令转化为机器人可执行的动作序列^[15]。同时,该系统还能从人类遥操作示范中学习底层技能,必要时由机器人自主完成某些步骤。这意味着未来的遥操作有望从“人直接控制每个关节”提升到“人赋予目标, AI 生成动作方案, 机器人自主完成细节”。生成式 AI 还能用于增强反馈和环境理解,例如,利用生成模型预测远程环境中未感知部分的情况,提前规划避障路径,减少通信时延影响等。

当然,目前生成式策略在遥操作中的应用仍处于早期探索阶段。如何确保生成模型输出动作的可靠性和安全性是重要挑战之一^[31]。然而,其巨大的潜力不容忽视:通过学习大量多样的人类动作分布,生成模型有望为机器人提供高度概括的运动技能库;通过借鉴语言、大模型的推理能力,远程机器人或许能理解更抽象的人类意图并据此采取行动^[41]。可以预见,随着计算能力和算法的发展,生成式 AI 将成为具身遥操作控制模式演进的关键推动力量,使远程机器人的自主智能达到新的高度^[11]。

3 系统架构与多模态反馈设计

遥操作系统由人侧操控端—通信链路—机器人从端 3 部分闭环构成。操作者通过 VR/动捕/外骨骼等接口产生输入,经由感知与映射将位置姿态重定向、关节一对一映射,并融合凝视、足压等多源信号以估计意图;机器人侧采用分层/全身控制(导纳/阻抗、平衡与接触管理、被动性与鲁棒稳定等)将高层目标下发至执行层,保障动态稳定与安全。在通信与实时性上,系统面向低时延高可靠(有线/光纤/5G),并通过预测显示、云中继、服务质量控制与自适应压

缩,对抗时延与丢包。为提升临场感(telepresence)与操控精度,体系提供多模态反馈——高清低时延视觉(立体/深度/红外/360°)、双向音频,以及从局部到全身的力/触觉回传。配套工程措施(标定与时延校准、

日志与健康监测、限位与急停)贯穿各环节,最终形成在复杂环境中实时、稳健且安全的人机闭环。遥操作系统架构如图2所示。

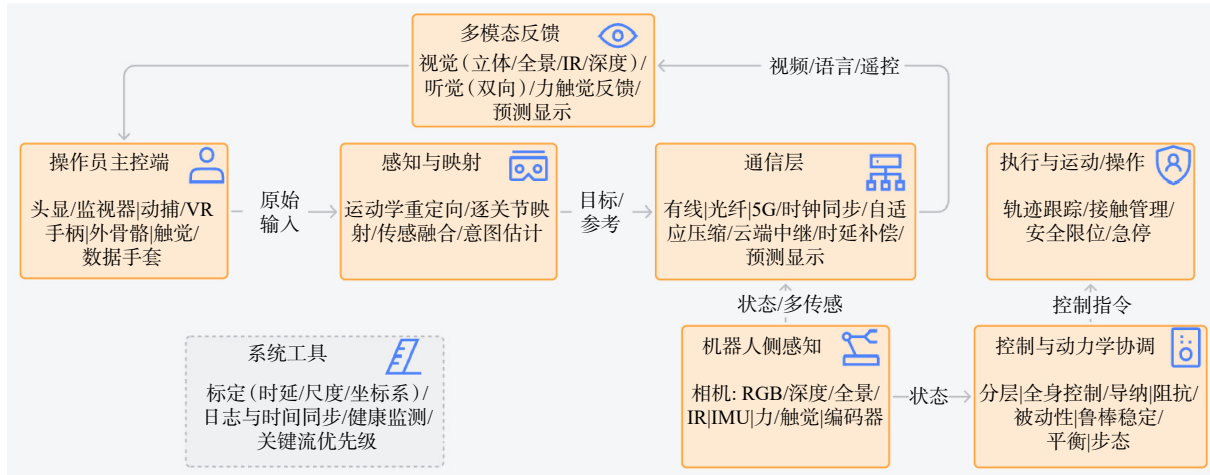


图2 遥操作系统架构

3.1 遥操作系统架构

典型的具身遥操作系统由人侧操控端、通信链路和机器人从端3大部分构成。在人侧,操作员通过前述各种接口设备(主控端)发出动作指令,并接收从机器人的环境感知和反馈信号。机器人侧则执行接收到的命令,并将自身状态和传感信息传回。两端通过有线或无线通信网络连接,形成闭环的人机系统^[3]。

在操作者一侧,系统需要将人类运动意图映射为机器人可执行的目标。针对不同接口,映射过程有所差异:对于位置/姿态类输入(如VR手柄、光学动捕),通常通过运动学重定向(retargeting)方法,将人体的位姿映射到机器人对应的关节空间^[15]。这涉及考虑人和机器人的尺寸比例、关节排列差异等。例如,人类手臂与机器人手臂长度不同时,系统可通过缩放和重定向方法进行校准,从而保证动作幅度的一致性。再如,若肩关节朝向不完全匹配,则可通过坐标系映射和姿态补偿避免误差,使动作不会因人机尺寸或结构差异而失真^[9,16,17,32]。

在机器人一侧,控制模块接收来自人侧的目标指令,并将其转化为底层电机控制信号。针对人形机器人这类高自由度、多接触的系统,常采用分层控制架构:上层根据人类指令生成机器人关节参考轨迹,下层采用反馈控制器跟踪参考并维持动态平衡等约

束^[3]。以HRP-2人形机器人为例,Dafarra等^[8]和Peer等^[42]采用导纳控制(admittance control)作为上肢双边遥操作的控制策略。在该方法中,远程机器人的关节参数由一个虚拟的弹簧-阻尼系统确定:当操作员通过主端施加力时,机器人关节产生相应位移,参数选取保证即使存在通信时延也可稳定地跟随人端操作。类似地,下肢的平衡控制通常并行进行,例如,利用力传感器和陀螺仪数据调整步态,以防止人在操作过程中因疏忽导致机器人失稳跌倒。全身双边控制目前仍在探索阶段,有研究尝试在操作员与机器人之间建立全身运动和受力的动态耦合关系^[14-15]。这需要机器人具备先进的本体感知(如关节力矩、足底压力)并实时反馈给人,使两者形成“如影随形”的耦合系统。虽然实时的全身力反馈尚未完全实现,但局部的力反馈已较为常见,例如,遥操作手部抓取时的握持力反馈、遇到障碍时的碰撞力反馈等^[11,17]。这些反馈通过主端设备(外骨骼、电动手套等)传递给操作员,提升了控制的稳定性和精细度。

遥操作对通信网络的要求极高,需要低延迟、高带宽和高可靠性。理想状态下,人发出的指令和机器人反馈应能实时同步,否则超过一定延迟会让操作者产生“时滞感”,甚至晕眩,严重时可能导致操作失败,甚至置于危险中^[43]。为降低时延影响,多数遥操作系

统采用有线或专用光纤通信线路;在无线方案中,新兴的 5G 网络以其毫秒级延迟被寄予厚望^[44-45]。为对抗时延与丢包,工程上常用 3 类手段:(1) 显示与交互侧的预测/渐进纠偏;(2) 控制侧的鲁棒稳定性设计(如被动性/约束控制);(3) 通信侧的自适应压缩与 QoS 策略(关键数据优先、带宽退化可用)^[46-48]。

通过上述感知、控制和通信模块的协同设计,现代遥操作系统能够在保障实时性的同时,实现复杂环境下人形机器人的稳定远程操控。在实际实现中,还涉及安全限位、紧急停止机制、系统延迟校准等大量工程细节,共同确保遥操作系统的鲁棒性和可靠性。

3.2 多模态反馈通道设计

在人机遥操作系统中,及时、真实的多模态反馈对于提升操作者的临场感至关重要。理想情况下,操作者应感受到如同亲临远程现场般的视觉、听觉和触觉体验,从而做出正确的判断和自然的动作调整^[3,44]。当前遥操作系统主要提供以下 3 种反馈通道。

视觉是人获取远程环境信息的最主要途径。通常机器人会安装立体相机或全景摄像头,将远程现场画面实时传输给操作者。操作者通过监视器或 VR 头盔看到机器人视角的 3D 场景,实现“所见即所控”。为增强沉浸感,不少系统采用第一人称视角,即将摄像头设置在人形机器人的头部位置并附带自由转动机构,使操作者转动自己的头部就能同步改变机器人摄像头朝向^[11]。Open-TeleVision 等工作进一步提出主动视觉反馈机制:根据操作者任务需要,机器人自主调整摄像头位置、焦距或显示参数,突出关键任务信息、滤除干扰,从而减轻操作者认知负担^[32]。除了可见光摄像头,有些遥操作场景还叠加了其他视觉通道,如红外热成像、深度相机、360°环视等,以增强环境感知^[12]。这些视觉数据往往通过屏幕上的多窗口或 MR 眼镜的画中画叠加呈现,确保操作者充分了解机器人的周边环境状态。总体而言,高清、低延迟的视觉反馈是遥操作沉浸体验的基础。

远程机器人的麦克风可拾取环境声音,如机械运行噪声、人声对话、报警声等,并通过音频流传送给操作者^[49]。声音反馈在某些任务中非常重要,例如,远程医疗诊断需要听诊器声音、工业遥操作需监听机器异响等。现代遥操作系统通常提供双向音频,使操作者不仅能听到远程声音,也可通过扬声器与远程人员交流,

实现自然的语言沟通。尽管听觉相对于视觉在机器人控制中的作用次要,但它显著提升了沉浸感和信息完备性。试想在灾区搜救遥操作中,操作者听到呼救声即可判断幸存者方位,而不仅仅依赖摄像头寻找^[13]。

力反馈是高端遥操作系统的核心特色之一。当远程机器人与环境发生物理交互时,系统会将接触力或运动阻力反馈给操作者,使其“感觉”到碰撞、重量、材质等信息^[9]。这通常通过带有制动器的主控设备(如有力反馈的操纵杆、外骨骼、电动触觉手套等)实现。例如,当机器人手臂抬起重物时,操作者外骨骼肩部会产生相应拉力;机器人手爪摸到粗糙表面时,操作者指尖的触觉手套振子会产生振动反馈^[16]。双边遥操作(bilateral teleoperation)利用力反馈让操作者与机器人形成闭环力耦合,大幅提升了操作的精确度和安全性。在人形机器人遥操作中,力反馈更具挑战性,因为涉及全身接触。现有系统多聚焦于上肢和手部,如握持力、碰撞力等;对于下肢平衡力和全身惯性力的反馈,目前大多以视觉替代或简单提醒(如座椅震动)实现^[20]。不过,iCub3 Avatar 系统已尝试重量反馈:当机器人举起重物时,外骨骼同步给操作者施加拉力,让其直观感知重量^[8]。可以预见,随着外骨骼和触觉设备的发展,未来会有更多维度的力触觉反馈融入遥操作,让人真正“身临其境”地操控机器人。

综合来看,多模态反馈的目标是在远程环境和操作者之间建立完整的信息回路,让操作者通过视觉“看到”、通过听觉“听到”、通过触觉“摸到”机器人所处的环境状态。这不仅增强了任务效率和准确性,也改善了操作者体验,使其长时间操作仍保持良好态势感知而不易疲劳。全沉浸式遥操作正朝着“欺骗”人类感官的方向发展,让操作者仿佛完全置身于远端环境。这种极致沉浸正是具身机器人遥操作最终追求的目标之一。在了解了遥操作系统的构成与反馈设计后,如何公平客观地评价不同系统的性能就成为一个关键问题,下文将介绍评估指标和基准测试。

4 性能评估方法与公开基准平台

评估遥操作系统需要综合考虑任务完成效果,可分为 4 类。(1) 任务性能:成功率、完成时间、轨迹误差、接触/装配精度等。(2) 安全性:碰撞/跌倒事件

率、力/速度越限次数、急停触发频次等。(3) 人因工程: 工作负荷、学习成本(上手时间)、疲劳指标(主观+生理)等。(4) 通信鲁棒性: 端到端时延阈值、丢包敏感性、降带宽退化曲线等。基于上述维度, 未来基准平台除任务集统一外, 还须统一日志格式、时钟同步/对齐方法与公开评测协议, 以支撑跨系统复现与工程选型。传统研究往往在自定义场景测试系统性能, 导致不同工作间缺乏直接可比性。近年来, 统一的基准测试平台开始出现, 为不同方案提供了公平比较的环境。例如, 2025 年提出的 TeleOpBench^[19]为针对双臂灵巧遥操作的大规模仿真基准平台。该平台提供了 30 个高保真任务场景(从基本抓取放置、工具使用到复杂双臂协作), 并评估了光学动捕、VR 控制器、上肢外骨骼和单目视觉 4 类典型接口。研究人员在此基准上进行测试, 结果表明仿真评测与真实机器人实验高度相关。验证了其作为统一度量标准的有效性。TeleOpBench 为不同遥操作方法的对比提供了“试金石”, 推动了性能评测的规范化。

除了客观的任务指标, 遥操作系统的人因评估同样重要。例如, 操作者的工作负荷、疲劳度和主观满意度常通过 NASA-TLX^[50]等问卷在实验后收集。这些主观指标反映了系统界面设计的合理性和长期使用的舒适性。如果一个系统任务成功率高, 但操作者极易疲劳或难度大, 则其实用性会受到限制。因此, 理想的遥操作系统应在保证任务性能的同时尽量降低操作者负担。一些研究还通过监测心率、肌电等生理信号量化操作者紧张度, 或统计训练时间和失效率评估系统上手难度。将这些人因数据与性能指标结合, 可刻画系统的优劣, 为系统设计提供改进方向。

5 开源平台与数据集概览

开放的共享平台和数据资源对于加速具身遥操作领域的发展起到了重要作用。多项前沿工作选择在发表的同时公开其代码、硬件设计乃至采集的数据。这一趋势极大降低了后来者重复实验和改进算法的门槛, 推动了社区协作。

在硬件平台方面, 低成本开源遥操作系统开始涌现。2025 年银河通用机器人与清华大学发布的 OpenWBT 是一款全开源的人形机器人全身遥操作系

统^[13]。OpenWBT 提供了从 VR 界面到机器人接口的完整软件栈, 新手也能在数小时内部署, 并支持多种型号的人形机器人。这种开源系统让更多研究团队能够搭建属于自己的遥操作平台, 进行针对性的研究, 而不必从零开始开发整套系统。另一个案例是 UIUC 的 CHILD 系统, 其硬件设计(包括 3D 打印模型、电路方案等)全部开源, 材料成本不到 1000 美元^[16]。CHILD 的开源意味着任何实验室都可以复现一个功能类似的便携式全身遥操作装置, 从而验证论文结果或在其基础上改进。OpenRobotLab 的 HOMIE 系统同样将设计细节和控制代码开源发布^[9]。HOMIE 通过廉价硬件实现了高效的半自主遥操作, 其开源代码为社区提供了宝贵参考。此外, 跨平台的视觉-外骨骼遥操作系统 ACE、双腕视觉引导遥操作 Bunny-VisionPro、沉浸式视觉反馈遥操作 Open-TeleVision、人体-机器人影子跟随系统 HumanPlus 等项目也均在 GitHub 上开放了源码。总的来看, 当今遥操作领域新的研究几乎都倾向于开放源码, 以期凝聚社区力量, 避免重复“造轮子”^[11,17,20]。

在数据资源方面, 上文提到的一些公开数据集正逐步为研究者所利用。例如, TeleOpBench 提供的全套模拟任务环境及评测工具, 其实质也是一个可重复生成遥操作交互数据的资源库。一些工作甚至专门收集和发布遥操作示范数据: OmniH2O-6 数据集涵盖 6 种日常任务的人类遥操作演示, 包括机器人行走、搬运、运动交互等多模态记录, 可用于训练和验证整身运动策略^[15,19]。再如 Iit 等机构在 Avatar XPrize 过程中累积了大量类人遥操作数据, 尽管未完全公开, 但也推动了后续研究对远程化身机器人的理解。随着众包遥操作理念出现, 未来可能由众多非专家用户贡献遥操作数据, 从而构建规模空前的教学数据集, 供机器人学习更普适的技能。

此外, 还有一些社区整合性资源在起作用。例如, 一些研究者整理了遥操作相关论文、代码、数据的网站(如 Ze 等^[30]维护的“人形机器人学习资料库”), 方便新人快速了解现有成果并获取资源。各大机器人挑战赛和开源项目的积累, 也让硬件厂商开始关注标准化接口和模块化组件, 为构建遥操作系统提供即插即用的部件。总体上, 开源平台与数据共享正在将具身遥操作带入一个良性循环: 更多人用得起、玩得

起这些系统,就会产生更多反馈和创意,进而改进系统性能、丰富数据,吸引更多人加入。在学术界和工业界的共同推动下,遥操作生态正在快速成熟^[8,29]。

6 发展趋势与挑战

面向未来,具身机器人遥操作技术将在沉浸交互、通信能力、智能化水平、标准体系与工程可及性等方面持续演进,同时面临安全性、稳定性与工程化落地等关键挑战。(1) 更高沉浸感的人机交互。融合多模态感知与反馈,提升临场感与态势感知,同时需兼顾系统稳定性与人因负荷。(2) 超低时延与广域通信。5G/6G 与卫星通信拓展遥操作应用边界,超远距离场景下时延、抖动与链路稳定性仍是关键工程约束,可通过预测显示、边缘计算与通信-控制协同设计缓解。(3) AI 驱动的人机协同与自主辅助。通过意图识别、运动预测与策略生成等方法提供路径优化、约束保护与误操作校正等辅助,遥操作积累的示教数据亦成为机器人自主技能学习的重要数据入口。(4) 标准化平台与评测体系建设。需从任务性能、安全性、人因工程与通信鲁棒性等维度建立系统化评测体系,通过模块化接口与统一协议提升互操作性。(5) 低成本化与工程可及性。采用消费级传感、简化接口设计及增强软件能力,在保障安全与性能前提下降低硬件成本与部署门槛。(6) 工程应用牵引与现实挑战。在核应急、高危化工、灾害救援、深海与空间站维护等高风险场景中具有不可替代作用,评测体系需重点关注高容错、安全守护、低带宽可用性与长时连续作业能力。

具身机器人遥操作正从单一远程操控向智能化、多样化、低成本化方向演进。在此基础上,归纳出 3 条规律性认识。(1) 降低人机负担是持续演进方向:从直接控制到共享控制,再到生成式与大模型辅助,核心目标是将操作者从低层动作细节中解放,转向高层意图表达。(2) 安全性与鲁棒性构成工程落地关键瓶颈:高自由度、多接触特性叠加通信时延与丢包风险,使安全守护与约束控制具有基础性地位。(3) 数据驱动的自主学习闭环日益凸显:遥操作正从“替代人工操作”扩展为“支撑自主能力演进”的数据基础设施。

展望未来,随着 AI 方法的持续引入与标准化体系的完善,遥操作系统在性能、可靠性与适用范围上有望进一步提升,在高风险作业、应急救援及日常服务等场景中将继续发挥连接人类智能与机器执行能力的重要作用。

参考文献(References)

- [1] Pfeifer R, Bongard J, Grand S, et al. How the body shapes the say we think[M]. Cambridge, Mass: MIT Press, 2007.
- [2] Toet A, Kuling I A, Krom B N, et al. Toward enhanced teleoperation through embodiment[J]. *Frontiers in Robotics and AI*, 2020, 7: 14.
- [3] Darvish K, Penco L, Ramos J, et al. Teleoperation of humanoid robots: A survey[J/OL]. 2023: arXiv: 2301.04317. <https://arxiv.org/abs/2301.04317>.
- [4] Luo R, Zolotas M, Moore D, et al. User-customizable shared control for robot teleoperation via virtual reality[J/OL]. 2024: arXiv: 2403.13177. <https://arxiv.org/abs/2403.13177>.
- [5] Zhang Q L, Liu Q H, Duan J G, et al. Research on teleoperated virtual reality human-robot five-dimensional collaboration system[J]. *Biomimetics*, 2023, 8(8): 605.
- [6] 兰津卜, 赵文博, 朱凯, 等. 基于具身智能的移动操作机器人系统发展研究[J]. *中国工程科学*, 2024, 26(1): 139-148.
- [7] Fang B, Sun F C, Liu H P, et al. Wearable technology for robotic manipulation and learning[M]. Singapore: Springer Singapore, 2020.
- [8] Dafarra S, Pattacini U, Romualdi G, et al. iCub3 avatar system: Enabling remote fully immersive embodiment of humanoid robots[J]. *Science Robotics*, 2024, 9(86): eadh3834.
- [9] Ben Q W, Jia F Y, Zeng J, et al. HOMIE: Humanoid locomanipulation with isomorphic exoskeleton cockpit[J/OL]. 2025: arXiv: 2502.13013. <https://arxiv.org/abs/2502.13013>.
- [10] Li Y X, Lin Y T, Cui J M, et al. CLONE: Closed-loop whole-body humanoid teleoperation for long-horizon tasks[J/OL]. 2025: arXiv: 2506.08931. <https://arxiv.org/abs/2506.08931>.
- [11] Ding R Y, Qin Y Z, Zhu J Y, et al. Bunny-VisionPro: Real-time bimanual dexterous teleoperation for imitation learning[J/OL]. 2024: arXiv: 2407.03162. <https://arxiv.org/abs/2407.03162>.
- [12] Lu C H, Cheng X X, Li J L, et al. Mobile-TeleVision: Predictive motion priors for humanoid whole-body control[J/OL]. 2024: arXiv: 2412.07773. <https://arxiv.org/abs/2412.07773>.
- [13] Zhang Z K, Chen C, Xue H, et al. Unleashing humanoid reaching potential via real-world-ready skill space[J/OL].

- 2025: arXiv: 2505.10918. <https://arxiv.org/abs/2505.10918>.
- [14] Ze Y J, Chen Z X, Araújo J P, et al. TWIST: Teleoperated whole-body imitation system[J/OL]. 2025: arXiv: 2505.02833. <https://arxiv.org/abs/2505.02833>.
- [15] He T R, Luo Z Y, He X L, et al. OmniH₂O: Universal and dexterous human-to-humanoid whole-body teleoperation and learning[J/OL]. 2024: arXiv: 2406.08858. <https://arxiv.org/abs/2406.08858>.
- [16] Myers N, Kwon O, Yamsani S, et al. CHILD (controller for humanoid imitation and live demonstration): A whole-body humanoid teleoperation system[J/OL]. 2025: arXiv: 2508.00162. <https://arxiv.org/abs/2508.00162>.
- [17] Yang S Q, Liu M H, Qin Y Z, et al. ACE: A cross-platform visual-exoskeletons system for low-cost dexterous teleoperation[J/OL]. 2024: arXiv: 2408.11805. <https://arxiv.org/abs/2408.11805>.
- [18] Walton D R, Steed A. Accurate real-time occlusion for mixed reality[C]//Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology. Gothenburg, Sweden: ACM, 2017.
- [19] Li H Y, Zhao Q, Xu H R, et al. TeleOpBench: A simulator-centric benchmark for dual-arm dexterous teleoperation[J/OL]. 2025: arXiv: 2505.12748. <https://arxiv.org/abs/2505.12748>.
- [20] Fu Z P, Zhao Q Q, Wu Q, et al. HumanPlus: Humanoid shadowing and imitation from humans[J/OL]. 2024: arXiv: 2406.10454. <https://arxiv.org/abs/2406.10454>.
- [21] He T R, Luo Z Y, Xiao W L, et al. Learning human-to-humanoid real-time whole-body teleoperation[J/OL]. 2024: arXiv: 2403.04436. <https://arxiv.org/abs/2403.04436>.
- [22] Cao Z, Hidalgo G, Simon T, et al. OpenPose: Realtime multi-person 2D pose estimation using part affinity fields[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(1): 172–186.
- [23] Zhang F, Bazarevsky V, Vakunov A, et al. MediaPipe hands: On-device real-time hand tracking[J/OL]. 2020: arXiv: 2006.10214. <https://arxiv.org/abs/2006.10214>.
- [24] Da Vinci surgical system white papers[EB/OL]. [2025–10–11]. <https://www.intuitive.com/en-us/healthcare-professionals/academics>.
- [25] DARPA robotics challenge finals[C]//Proceedings of the DARPA Robotics Challenge Finals. Pomona, California: DARPA, 2015.
- [26] Jeon H J, Losey D P, Sadigh D. Shared autonomy with learned latent actions[J/OL]. 2020: arXiv:2005.03210.
- [27] Xie B W, Han M J, Jin J, et al. A generative model-based predictive display for robotic teleoperation[C]//Proceedings of IEEE International Conference on Robotics and Automation (ICRA). Piscataway, NJ: IEEE, 2021: 2407–2413.
- [28] Gonzalez G, Agarwal M, Balakuntala M V, et al. DESERTS: Delay-tolerant SEmi-autonomous robot teleoperation for surgery[C]//Proceedings of IEEE International Conference on Robotics and Automation (ICRA). Piscataway, NJ: IEEE, 2021: 12693–12700.
- [29] Seo M, Han S, Sim K, et al. Deep imitation learning for humanoid loco-manipulation through human teleoperation[J/OL]. 2023: arXiv: 2309.01952. <https://arxiv.org/abs/2309.01952>.
- [30] Ze Y J, Chen Z X, Wang W H, et al. Generalizable humanoid manipulation with 3D diffusion policies[J/OL]. 2024: arXiv: 2410.10803. <https://arxiv.org/abs/2410.10803>.
- [31] Chi C, Xu Z J, Feng S Y, et al. Diffusion policy: Visuomotor policy learning via action diffusion[J/OL]. 2023: arXiv: 2303.04137. <https://arxiv.org/abs/2303.04137>.
- [32] Cheng X X, Li J L, Yang S Q, et al. Open-TeleVision: Teleoperation with immersive active visual feedback[J/OL]. 2024: arXiv: 2407.01512. <https://arxiv.org/abs/2407.01512>.
- [33] Taylor R H. A perspective on medical robotics[J]. *Proceedings of the IEEE*, 2006, 94(9): 1652–1664.
- [34] Sheridan T B. Telerobotics, automation, and human supervisory control[M]. Cambridge, MA: MIT Press, 2003.
- [35] Nagatani K, Kiribayashi S, Okada Y, et al. Emergency response to the nuclear accident at the Fukushima Daiichi Nuclear Power Plants using mobile rescue robots[J]. *Journal of Field Robotics*, 2013, 30(1): 44–63.
- [36] Rakita D, Mutlu B, Gleicher M. RelaxedIK: Real-time synthesis of accurate and feasible robot arm motion[C]//Proceedings of Robotics: Science and Systems XIV. Robotics: Science and Systems Foundation. Pittsburgh: IEEE, 2018: 26–30.
- [37] Rahal R. Shared control and authority distribution for robotic teleoperation[D]. Rennes: Université Rennes 1, 2020.
- [38] Hussein A, Gaber M M, Elyan E, et al. Imitation learning: A survey of learning methods[J]. *ACM Computing Surveys*, 2018, 50(2): 1–35.
- [39] Schaal S. Is imitation learning the route to humanoid robots?[J]. *Trends in Cognitive Sciences*, 1999, 3(6): 233–242.
- [40] Ross S, Gordon G J, Bagnell J A. A reduction of imitation learning and structured prediction to no-regret online learning[C]//Proceedings of the 14th International Conference on Artificial Intelligence and Statistics. New York: PMLR, 2011, 15: 627–635.
- [41] Brohan A, Brown N, Carbajal J, et al. RT-2: Vision-

- language-action models transfer web knowledge to robotic control[J/OL]. 2023: arXiv: 2307.15818. <https://arxiv.org/abs/2307.15818>.
- [42] Peer A, Hirche S, Weber C, et al. Intercontinental cooperative telemanipulation between Germany and Japan[C]// Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, NJ: IEEE, 2008: 2715–2716.
- [43] Kamtam S B, Lu Q, Bouali F, et al. Network latency in teleoperation of connected and autonomous vehicles: A review of trends, challenges, and mitigation strategies[J]. *Sensors*, 2024, 24(12): 3957.
- [44] Farajiparvar P, Ying H, Pandya A. A brief survey of telerobotic time delay mitigation[J]. *Frontiers in Robotics and AI*, 2020, 7: 578805.
- [45] Kizilkaya B, Popoola O, Zhao G, et al. 5G-based low-latency teleoperation: two-way timeout approach[C]// Towards autonomous robotic systems. Cham: Springer Nature Switzerland, 2023.
- [46] Xi L, Zhang H C, Liu J R, et al. A novel P+d control scheme for time-delayed telerobotic systems with damping adjustment[J]. *Processes*, 2025, 13(3): 611.
- [47] Zhu H J, Sharma M, Pfeiffer K, et al. Enabling remote whole-body control with 5G edge computing[J/OL]. 2020: arXiv: 2008.08243. <https://arxiv.org/abs/2008.08243>.
- [48] Kizilkaya B, She C Y, Zhao G D, et al. Task-oriented prediction and communication co-design for haptic communications[J/OL]. 2023: arXiv: 2302.11064. <https://arxiv.org/abs/2302.11064>.
- [49] Purushottam A, Yan J, Yu C, et al. Heavy lifting tasks via haptic teleoperation of a wheeled humanoid[J/OL]. 2025: arXiv: 2505.19530. <https://arxiv.org/abs/2505.19530>.
- [50] Hart S G, Staveland L E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research[M]// Human Mental Workload. Amsterdam: Elsevier, 1988: 139–183.

A survey on embodied robot teleoperation

NI Ziqiang, XIE Shaoxuan, LIU Xuecheng, YANG Xiang, LIU You, YAO Guocai*

Beijing Academy of Artificial Intelligence, Beijing 100089, China

Abstract Embodied robot teleoperation will remain indispensable before robots can achieve fully human-level autonomy. This paper first categorizes human-machine interaction modalities for teleoperation, including exoskeletons, virtual/mixed reality, motion capture/inertial sensing, and vision-based interfaces. It then analyzes the evolution of teleoperation control paradigms, covering direct control, shared control, imitation-learning-based approaches, and the recent introduction of generative strategies, highlighting their underlying principles and application domains. Next, representative system architectures and technical implementations are introduced, encompassing perception and mapping, control algorithms, and communication modules, while discussing the role of multimodal feedback—such as vision, force, and touch—in enhancing immersion. Furthermore, performance evaluation methods and the latest benchmarks are reviewed, emphasizing the significance of open-source software/hardware platforms and data resources in advancing the field. Finally, the paper summarizes key challenges and future directions, including improving the level of intelligence, reducing costs and barriers to adoption, and establishing standardized frameworks.

Keywords embodied robots; teleoperation; human-machine interaction; control paradigms ●



(责任编辑 王微)