DOI 10.15302/J-SSCAE-2024.01.010

基于具身智能的移动操作机器人系统发展研究

Lan Fengbo，赵文博[1]，朱凯[1]，张涛1,2*

（1.清华大学自动化系，北京100084;2.北京信息科学与技术国家研究中心，北京100084）

摘要：具身智能是新一轮科技革命和产业变革中的战略性技术，是当前世界各国重点竞争的前沿高地之一；移动操作机器人系统因其优秀的运动、规划、执行能力成为具身技术首选的硬件载体；基于具身智能的移动操作机器人系统作为实现跨领域、多场景、多功能的自主具身智能平台，将成为引领未来新一代信息技术和人工智能发展的关键。本文从基于具身智能的移动操作机器人系统发展的需求出发，总结了基于具身智能的移动操作机器人系统的发展现状，分析了该领域发展面临的问题和挑战，提出了涵盖多模态感知技术、世界认知与理解技术、智能自主决策技术、运动与操作联合规划技术等基于具身智能的移动操作机器人系统的共性关键技术。基于此，本文从国家政策倾斜、共性技术突破、交叉学科建设与人才培养、综合验证平台构建等方面提出了对策建议，以期助力具身智能发展浪潮下我国移动操作机器人领域的长足发展。

关键词：具身智能；移动操作机器人；任务和运动联合规划；智能决策

中图分类号：TP242.6　　**文献标识码：**A

Development of Mobile Manipulator Robot System with Embodied Intelligence

Lan Fengbo[1],Zhao Wenbo[1],Zhu Kai[1],Zhang Tao1,2*

(1.Department of Automation,Tsinghua University,Beijing 100084,China;2.Beijing National Research Center for Information Science and Technology,Beijing 100084,China)

Abstract:Embodied intelligence stands as a strategic technology in the ongoing scientific and technological revolution,forming a frontier in global competition.The mobile manipulator robot system,with its exceptional mobility,planning,and execution capabilities,has become the preferred hardware carrier for embodied intelligence.Moreover,the mobile manipulator robot system, rooted in embodied intelligence,emerges as a pivotal platform capable of cross-domain functionality.Positioned at the forefront of a new era in information technology and artificial intelligence,this system is integral for future development. Addressing the strategic demand for embodied-intelligence-based mobile manipulator robot systems,this study presents an overview of the current developmental landscape.It delves into the challenges faced by this field,proposing key common technologies such as multimodal perception,world cognition,intelligent autonomous decision-making,and joint planning for movement and manipulation.Furthermore,the study offers recommendations for advancing the field,encompassing national policy support,breakthroughs in common technologies, interdisciplinary collaboration,talent cultivation,and the construction of comprehensive verification platforms.These suggestions aim to facilitate the rapid progress of mobile manipulator robots in China amid the wave of embodied intelligence development.

Keywords: embodied intelligence;mobile manipulator robot;joint planning for movement and manipulation;intelligent decision-making

收稿日期：2023-11-09；**修回日期：**2023-12-15

通讯作者：*张涛，清华大学自动化系教授，研究方向为机器人与人工智能；　E-mail:taozhang@tsinghua.edu.cn

资助项目：中国工程院咨询项目“新一代人工智能及产业集群发展战略研究”（2022-PP-07）

本刊网址：www.engineering.org.cn/chljournal/sscae

公众号 · 智能机器人科技

一、前言

人工智能（AI） 是引领新一轮科技革命与产业 变革的战略性技术之一，对助力我国现代化产业体系建设、推动战略性新兴产业融合集群发展、赋能 各行各业发展具有重要意义。随着通用AI技术的 加速突破，技术基础扎实且可多场景适用的移动操 作机器人作为具身智能的优质载体，在全球科技界 与产业界掀起了新一轮的研究浪潮，成为新一轮科技革命中的前沿技术高地之一。

基于具身智能的移动操作机器人系统（简称 “具身智能移动操作机器人”）旨在构建具备自主 环境感知、充分理解认知、流畅人机交互、可靠智 能决策与自然运动操作规划的机器人系统，依托跨 领域、多场景、多功能的自主具身智能平台，为传 统移动操作机器人升级赋能，引领未来移动操作机 器人的行业发展。在具备可感知、理解与决策的类 脑结构后，移动操作机器人能够自主理解并完成人 类下达的高级指令，实现真正的通用智能。

与传统移动机器人相比，具身智能移动操作机器人能够完成一些通常需要人类智慧才能完成的复杂工作，随着其技术不断发展与成熟，必将给人类社会带来革命性的变革。具身智能移动操作机器人在服务、餐饮、医疗、智能家居、无人配送等民用领域，智能工厂、智能制造等工业领域以及单兵作战等军事领域，都拥有广阔的应用前景。目前，国内外对于具身智能移动操作机器人的研发大多仍处于实验室测试阶段，针对特定场景、特定任务的具身智能移动操作机器人虽然有了较大的发展，但整体技术并不成熟，尚未实现产业化与商品化。学术界对于移动操作机器人的研究主要围绕环境感知、运动控制、路径规划、车臂协同等方面展开，对具身智能技术和移动操作机器人技术的研究处于平行发展阶段。本文从具身智能移动操作机器人发展的需求出发，厘清具身智能移动操作机器人的发展现状，分析面临的挑战，总结相关关键技术，并提出发展建议，以期为具身智能移动操作机器人研究提供参考。

二、基于具身智能的移动操作机器人系统发展现状

移动操作机器人泛指具备移动与抓取操作能力的机器人，基本结构由移动基座、机械臂、操作末端共同组成，进而发展出类人结构。移动操作机器人拥有形态优势和移动操作能力，是最接近人类形态的机器人构型，是具身智能最理想的硬件载体。移动操作机器人技术发展历史悠久，经过长足发展，已形成相对成熟的技术体系。具身智能作为通用AI 技术的重要实现手段，近年来正在不断取得全新突破。具身智能移动操作机器人呈现出广阔的应用前景。

（一）移动操作机器人

移动操作机器人需要在未知环境中完成感知、导航与控制等任务，主要由移动底盘、机械臂和操作末端三部分组成，核心技术包括感知、导航与控制技术，灵巧操作控制技术等。移动操作机器人技术的发展使机器人能够更精确、快速、稳定地利用多模态数据来感知周围环境，进行更高效的运动控制和路径规划。随着深度学习技术的发展和应用，基于深度学习和强化学习的机器人控制技术、多模态感知技术将得到进一步发展，推动移动操作机器人的感知、规划、控制能力进一步提升。

移动操作机器人在感知、导航与控制技术方面的发展现状如下。①在感知技术层面，移动操作机器人依靠传感器来感知外部的物理世界，通过装配相机、雷达、超声波、红外等感知传感器以及惯性测量单元、编码器等传感器来确定自身位置姿态和运动状态。通过多传感器融合技术，移动机器人可以利用多源感知数据来提高自身感知的精度和鲁棒性，实现对环境的高精度实时感知。此外，感知层还需要完成对未知环境的建图任务。定位与建图技术用于解决移动机器人在真实物理世界中的环境重建和自身定位问题，是移动机器人的关键技术之一。目前，定位与建图技术研究通常采用激光测距仪、计算机视觉两种方法3：前者构建的点云地图 精度较高，算法相对简单，在光照不足、明暗变化 大的环境下鲁棒性强4，但是难以展现较好的环境 细节信息；后者得到的视觉图像包含更加丰富的环 境特征信息，使机器人能够在更大范围内完成任 务，但是算法设计复杂，对光照等条件要求比较苛 刻。②在导航技术层面，移动操作机器人根据目标 点和感知地图，实时生成离散航路点序列或连续期望轨迹。移动操作机器人导航技术主要包括

全局路径规划、局部避碰与避障等技术。移动操作机器人导航任务主要包括以地图的形式生成世界模型，计算从起始位置到目标位置的无碰撞轨迹，沿着计算的轨迹移动，避免与障碍物碰撞6。③在控制技术层面，移动操作机器人不仅需要具备移动机器人的感知、定位与导航运动能力，还需要具备精确控制移动机器人抵达预期航路点的能力。目前，机器人控制技术相对成熟，通过对机器人建立运动学和动力学模型，采用基于全局线性化的控制、基于近似线性化的控制、基于李雅普诺夫理论的控制等方法，实现较为稳定的控制。经典的机器人控制策略有计算力矩控制、鲁棒控制、滑模控制、自适应控制、神经网络控制、模糊逻辑控制、自抗扰控制与柔性控制5。

在灵巧操作控制技术方面，传统机器人的操作末端通常为夹爪或夹手，缺乏机动性与灵巧性；移动操作机器人将具备更加灵巧通用的操作末端，如类手结构、柔顺捕获结构等，以提升末端抓取的通用性，拥有任意物体抓取、工具使用、柔性物体操作等高级技能，进而可以在工业环境中完成 装配、焊接、搬运等任务以及在家庭场景中完成厨具使用、家庭清洁等工作。除此之外，计算机视觉技术可以帮助移动操作机器人与周围环境进行互动，增强其智能化和自主性，使其像人一样感知、理解和响应环境，对提升机器人视觉跟踪、物体识别、移动抓取、人机交互等能力都有重要意义。机械臂控制技术和灵巧操作技术助力移动操作机器人完成高精度、高性能的复杂任务。移动操作机器人可以通过相机等视觉传感器、压敏阵列等触觉提升智能操作、灵巧操作和交互操作能力。受限于硬件资源与工作场景，移动操作机器人主要通过图像数据强化其操作能力，利用计算机视觉技术输入相机等传感器的图像像素并转换为环境中物体的类别、位置、姿态、速度以及人类的面部表情、手势等具体信息，运用计算机算法和机器学习算法对数字图像进行分析、处理、识别和理解，进而实现智能化应用。

得益于其感知、移动、抓取和灵巧操作等能力的提升，移动操作机器人在民用和军用等领域均拥有广阔的应用前景和应用价值。①在民用领域，移动操作机器人在智慧医疗、智能工厂、交通物流、实验室助手、家居服务、餐饮服务、住宿服务等行业涌现出多种落地产品[12]，为智慧社会构建提供了 强大推力。在医疗方面，移动操作机器人可以完成 药品和工具递送、病人看护、辅助诊疗等任务。在 制造业中，移动操作机器人可以实现传统机械臂无 法实现的功能，提升工业生产的自主化水平。在实 验室助手方面，移动操作机器人可以帮助实验人员 开展化学实验工作[13]，如中国科学技术大学集成移 动机器人、化学工作站、智能操作系统、科学数据 库，研制出数据智能驱动的全流程机器“化学家”。 在家居服务方面，移动操作机器人可以完成物品递 送、开门关门、垃圾清理等便捷居民生活的任务。 在餐饮和住宿服务方面，智能移动机器人可以辅助 完成上餐、餐具回收、物品递送等任务。②在军用 领域，移动操作机器人可以辅助作战部队开展侦 察、跨越险阻、弹药运输、后勤保障等任务，提高 作战效率和丰富作战手段。具备一定或较高智能的 移动操作机器人的广泛应用，为具身智能移动操作 机器人产业化提供了需求与应用场景。

（二）具身智能移动操作机器人

具身智能概念在1950年首次被提出，指能够与环境交互感知，具备自主规划、决策、行动、执行能力的机器人或虚拟环境中的仿真人，是AI 的终极形态。具身智能具备自主感知、认知、理解、推理、行动等类人或高于人类的能力，具备由“大脑”“小脑”“脑干”组成的完整人脑结构以及可实现行动的机器身体，如图1所示。其中，“大脑”部分负责对场景进行认知、理解、推断、分析，“小脑”部分开展柔性控制、协同控制、交互控制、灵巧操作等技能学习，“脑干”部分则承担能源调配、环境感知、信号分析与计算等工作。随着时代的发展和科技的进步，具身智能的概念被不断迭代和更新，尚未形成定论，但通识的理解是具有身体的智能技术，即为智能技术赋予行动能力。具身智能的实现依赖于AI 领域的计算机视觉技术、多模态感知融合技术、自然语言处理技术、因果推断技术、移动导航与规划技术等。区别于离线智能，具身智能要求机器人具备高度自主化且可进行感知、理解与决策的“脑”，具有稳定、安全、运动自然的机器人“身体”。此外，机器人还要具备在执行人类高级指令及与环境交互过程中进行在线学习、更新“脑”和“身体”的能力。

公众号 · 智能机器人科技

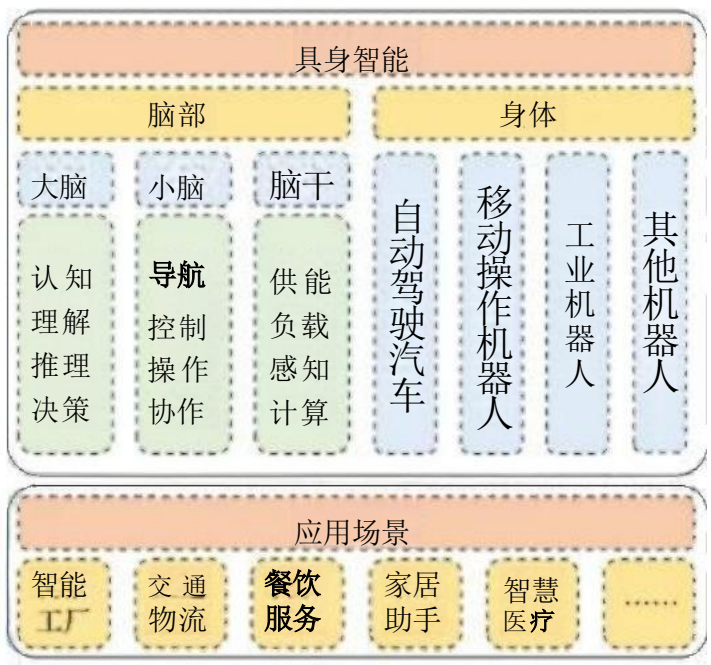


图 1 具身智能的内涵及其应用场景

2022年，OpenAI 公司研发的大语言模型ChatGPT 1.5被视为通用AI 技术领域取得的重要突破，进而使得具身智能被重新提起。大规模网络模 型基于海量优质数据训练后，在个性化应答、机器 翻译、语言理解、图像理解等方面的性能获得显著 提升，应用价值大幅提高。以大模型为内核，移动 操作机器人可以具备思考、理解、认知的能力。目 前，全球已出现多个成功商品化的大模型，国内包括百度公司研发的文心一言1.6、阿里巴巴公司研发的通义千问7等；国外包括OpenAI公司最新研发的GPT-4和图像生成大模型DALL·E 3.18、谷歌公司研发的Gemini 1.9、微软公司研发的数学推理大模型WizardMath 1.20、谷歌与柏林工业大学共同推出的视觉语言模型PaLM-E²等。这些大模型在常识推理、 代码完善、知识迁移等领域展现出较强的应用能 力。在语言识别与理解方面，基于大模型的聊天机 器人可以和人自由交流，完成资料检索、总结归 纳、日程安排、出行规划、行为规划等任务，具备 强大的文本认知与理解能力；针对图像信息，大模 型可以准确识别其中包含的各类物品，实现像素级 细粒度分割，并具备初步的空间识别能力；在点云 信息处理方面，大模型能够准确分割空间中的点云 区域，根据点云实现零训练的分类任务；在数学推 理、代码生成方面，部分专用大模型已经可以完成 定理证明、数学推导与计算，并可以自主编程实现 相应任务。目前，部分综合大模型已经具备较强的

文本、图像、点云的信息关联能力，能够根据文本信息准确地识别、分割或生成所需的图像、点云模型，并且支持将真实世界的连续传感器模态直接结合到语言模型中，从而建立单词文本和感知之间的联系。在近期的研究中（1221），综合大模型已经可以对人类高级命令生成分步指令。例如，由人类给出整 理房间的指令，大模型可以根据当前的图片信息判 断各类物品应摆放于何处，进而生成分布执行指令 完成该任务。

与传统移动操作机器人相比，具身智能移动操作机器人最大的进步在于其具备智慧大脑，能够实现对世界的认知理解进而进行智能决策。具身智能发展的核心是机器人“大脑”能力的进步。当前的具身智能发展主要集中在移动操作机器人“大脑”层面的研究。随着大模型技术尤其是多模态综合大模型技术的高速发展，具身智能移动操作机器人将具备更强大的思考、感知、认知、决策能力，实现更加通用自主的机器智能，利用多模态数据进行自我学习、自我适应、自我优化，自主筛选出最优的行动策略与最佳的解决方案以执行任务。

随着机器人技术和AI 技术的快速发展，移动操作机器人实物系统不断落地，出现了多个具备感知、导航、操作能力的移动操作机器人系统，如北京字节跳动科技有限公司的端到端机器人操作模型GR-11231、斯坦福大学Mobile ALOHA机器人24、谷歌DeepMind公司的视觉-文本多模态大模型RT-2以及最新推出的AutoRT[25]等。沈阳新松机器人自动化股份有限公司于2021年推出了全新的多可移动协作机器人，可搭载自主研发的视觉和夹具等执行单元，实现对物料的搬运、装配、检测和精密加工等功能应用（126）。杭州迦智科技有限公司推出了新款复合作业机器人MORA300，具有自主充电、自主定位导航、智能路径规划、第三方设施对接交互等功能（1271）。库卡机器人有限公司 (KUKA) 近年来推出了KMR iisy、KMR iiwa等移动协作机器人平台，可以与人一起合作，并以毫米级精度在工件上作业。此外，KUKA 公司的完全自主作业式移动机器人平台不需要电感线圈、地面标记或磁铁（I281）。 2023年7月，谷歌DeepMind 公司推出的机器人模型RT-2是一个全新的视觉－语言－动作模型，可以通过网络和机器人的数据进行学习，并将这些知识转化为机器人控制的通用指令（129）。2023年11月，One X公司与OpenAI 公司深度合作，开发了一款具身智能类人机器人EVE，可以实现对人类日常工作

环境的认知理解，在与环境交互的过程中学习、纠正、收集数据，完成自主居家、办公帮手任务30。波士顿动力公司研发了Spot机器狗，能够实现敏捷快速的探索与运动，支持在机器狗上扩展机械臂等多种外设；可以集成更多的传感器，提升通信和计算能力；通过元学习方式支持自主导航与探索，实现物理世界交互与无边界探索3。

总体而言，具身智能移动操作机器人拥有支持感觉和运动的机器身体，增强了主动感知与灵活执行的能力；不仅具备传统移动操作机器人的功能，还能够听懂人类语言，感知并理解外部环境，并据此分解任务、规划子任务，在移动中识别物体，与物理环境交互，最终完成相应任务。具身智能移动操作机器人正朝着形态多样化、功能全面化、任务通用化、行为自主化、交互人性化的方向发展，不断涌现出新的移动操作机器人设计思路和功能任务。移动操作机器人平台和实物系统的产业化进程不断加快，但其对外部环境和人类指令的理解和认知能力还有待提升，仍需人来发送具体的指令。未来，具身智能和移动操作机器人系统深入结合与落地，将推动移动操作机器人产业的进一步发展。

三、基于具身智能的移动操作机器人系统关键技术

当前，具身智能移动操作机器人发展形成的关键技术包括多模态感知技术、世界认知与理解技术、智能自主决策技术、运动与操作联合规划技术（见图2），旨在推动移动操作机器人系统的发展。

（一）多模态感知技术

多模态感知技术能够使具身智能移动操作机器人实现更高的自主性、高效性、通用性，增强对周围环境的局部感知能力，并为机器人提供丰富、稳定、准确的环境数据。室内复杂场景的运动感知信息往往存在多源性、异构性、动态性等特征，同时机器人需要面对环境光强变化、非全局特征感知、遮挡复原与推断等问题。对此，机器人可通过多角度图像信号、激光雷达（LiDAR）等不同模态信息数据关联以及多源图像数据融合，完成对周围局部

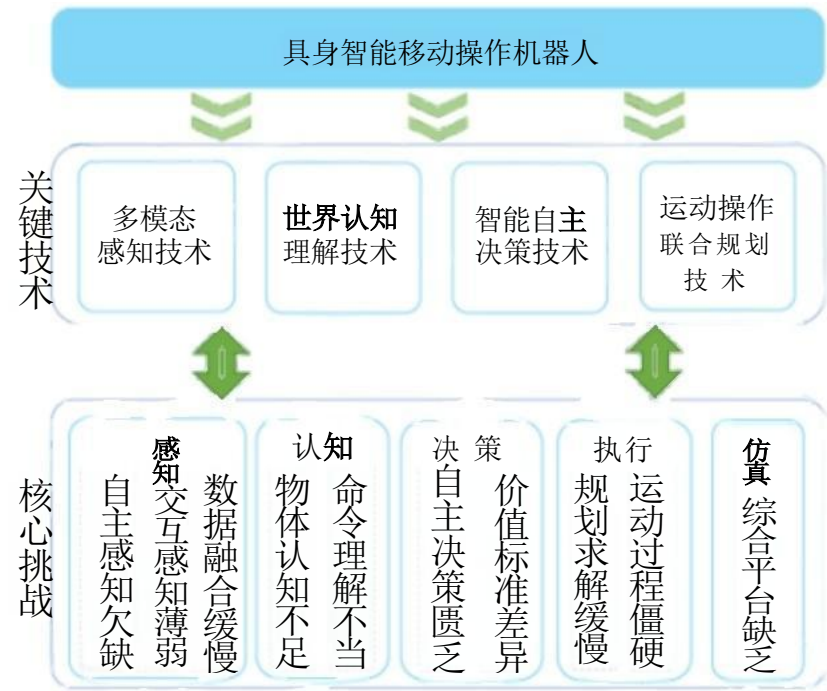


图 2　具身智能移动操作机器人发展与研究框架

环境的立体重建。针对室外开阔场景与复杂高噪环境，通过图像、LiDAR、热成像检测、全球定位系统等多源多模态传感器的多元融合技术，实现多传感器的优势互补，确保感知信息完备有效。通过目标检测、目标分割等技术，构建感知环境与实际环境的空间映射关系，形成多模态环境数据“一站式”融合智能处理系统，实现对局部环境空间、物体的即时感知与虚拟重建，为移动操作机器人系统对环境的认知与理解提供可靠数据来源。

（二）世界认知与理解技术

具身智能移动操作机器人与传统移动操作机器人相比，具备能够自主感知、认知理解、任务规划的智慧大脑。具身智能移动操作机器人构建感知世界与现实世界理解认知的共性关键技术发展方向主要有两种，一是基于深度学习方法，构建大模型，以环境感知数据为输入，通过迭代训练，形成对感知世界的经验认知与理解；二是通过对物理世界的物体认知，对物体运动、物体形变、工具使用等客观现象进行物理仿真和建模分析，构建世界共性物理模型，实现对局部环境的认知与理解。移动操作机器人系统利用智能技术完成对环境的感知与理解，进而对人类指令进行分析以构建上层任务的解析与规划，最终形成基于自然感知环境数据、可以独立自主对人类命令进行分析、实现任务分解的上层规划系统。

公众号 · 智能机器人科技

（三）智能自主决策技术

鲁棒、安全、最优的智能决策系统是具身智能移动操作机器人与环境和人类稳定交互、可靠执行、可行决策的关键。通过智能化技术，移动操作机器人具备自主决策与社交能力，可以通过自主生成、人机交互、机器人间交互三种方式生成决策。决策系统包含局部世界环境感知与物体认知理解的相互映射匹配和对齐、人类与机器人指令交互的双向价值对齐、人类命令与机器人可执行指令对齐三个主体部分。具体来说，具身智能移动操作机器人在获得人类高级命令后，融合移动操作机器人系统环境感知数据与物体认知理解，构建局部环境空间映射和物体种类、用途、使用方式的关联关系。在此基础上，决策系统进一步生成人类分步指令集，完成上游命令的任务分解，在人机交互过程中以在线学习方式动态进行双向人机价值对齐3，保障分 步决策符合人类价值体系。最终，构建人类命令与 机器人可执行指令的转化系统，完成分步决策结果 到移动操作机器人可执行指令的对齐转换，形成与 环境和人类稳定交互、自 主决策的移动操作机器人 决策系统。与一般智能机器人相比，该系统通常具 备社交导航、物体导航、人机协作、多机协作等高 级能力。现有的机器人自主决策研究主要有两种方 式，一种是以大语言模型为核心，通过对人类指令 进行预编码解码分析，设计输入信息，利用大语言模型给出决策方案以实现自主决策；另一种则是 通过对人类行为与物理世界进行解析分析，理解人 类行为与价值评判，进而根据当前感知信息推断规 划生成最优决策133。这两种方式各有优劣，目前由 于大语言模型技术的快速发展，第一种方法在近年 来更受关注。

（四）运动与操作联合规划技术

随着移动机器人技术的发展，移动操作机器人的种类不断增多，机器人的移动导航相关技术日趋成熟，机械臂的规划控制鲁棒性、自抗扰性、柔性控制技术日益完善。然而，移动操作机器人系 统作为具身智能的实际载体，其运动与操作联合规划技术尚未完备，致使具身智能的部署与应用受到较大掣肘。具身智能移动操作机器人已经具备对上游高级命令进行自动感知、自主决策、多机协作、人机交互的能力，而这对导航运动与操作的协同规划提出了更高的要求，单一的导航运动与操作 控制无法满足其灵巧性、高效性、连贯性、稳定 性、安全性的需求。具身智能移动操作机器人的运动与操作联合规划技术需要具备移动基座与机械臂的协同规划控制能力、多机协作与人机协作的工作能力。在此基础上，具身智能移动操作机器人能够 在复杂空间约束环境下完成典型任务，包括在室内 外复杂约束地形与环境下的移动、抓取、搜索、运 输、交互，融合环境感知数据进行局部建图与定 位、多关节路径规划、协同交互式操作、工具理解 与使用等；能够与其他机器人和人类并行统一协作 或串行交互传递完成难以独立完成的、更为复杂的 高级任务。此外，在共享同一工作空间的典型社会 化人机交互场景中，由于机器人的安全性和性能与 人类的运动自然耦合，机器人需要对人类未来的运 动进行实时推断，规划出安全、高效、符合社会规 范的路径，让人类长期接受机器人作为合作者，实现社交导航[35]。未来，移动操作机器人均需要在动 态社会化环境中进行安全的自主导航，人类也需要 很好地理解移动操作机器人的行为以便对其未来的 行为做出预测。人机交互共融技术将逐渐成为移动 操作机器人领域的研究重点。

四、基于具身智能的移动操作机器人系统发展面临的问题与挑战

当前，对机器人自主化、智能化的要求不断提高，传统离线智能模式的局限逐渐增多，已不能满足国家的战略需求与行业应用需要。随着多模态感知、AI、人机交互、自然语言处理、任务和运动规划等技术的不断发展，现有的具身智能移动操作机器人关键技术仍有待演进提升，具身智能移动操作机器人发展依然面临诸多问题与挑战。

（一）感知方面

在感知方面，具身智能移动操作机器人主要面临如下问题。一是自主感知能力欠缺。具身智能移动操作机器人不再局限于接受指令后仅对当前所在位置进行局部感知，而能够在复杂场景下根据人类提供的高级指令，自主判断对当前环境的感知程度，并进 一 步通过自主规划运动，实现动态感知，丰富自身对环境的感知程度，因此对其自主感知能

公众号 · 智能机器人科技

力提出了更高的要求。二是交互感知能力薄弱。具身智能移动操作机器人不再单独依靠人类低级指令进行规划执行，而是通过人类高级指令进行感知规划，完善对当前局部环境的多粒度探索，丰富环境感知数据。目前，移动操作机器人在约束环境下的自主感知技术仍处于发展阶段，没有有效的感知手段，对人机交互感知的准确性、即时性、有效性都提出了挑战。三是多模态数据融合与局部环境构建缓慢。移动操作机器人需要对周围局部环境进行充分完备的感知与重建，需要实现多模态数据融合与环境空间的三维映射，重建局部三维地图；具身智能移动操作机器人需要具备快速响应、自主规划能力，而当前技术在效率上无法保证即时高效。

（二）认知与理解方面

在认知与理解方面，具身智能移动操作机器人主要面临如下问题。 一是对环境物体的形态、功能、使用、交互方式存在认知欠缺。在具身智能下，移动操作机器人需要基于环境感知数据，完成对环境中各类物体的用途与使用方式等的认知和推断。当前，AI 领域的大模型技术虽然对常见物品具备一定的通识理解能力，但和人类经验之间仍存在偏差，无法构建环境中物体的关联关系，也难以根据环境中物体的实际形态、状态实现信息融合。这对移动操作机器人的决策生成与规划执行带来了可行性挑战。二是对人机交互的高级命令存在理解不足。在面对人类给出的诸如整理房间、完成仓库中货物运输、搜救被困人员等高级指令时，现有以大语言模型为基础的AI 技术尚无法通过与感知环境数据融合以给出合理的认知与任务分解，容易陷入经验认知陷阱，缺乏交互学习与矫正能力。

（三）决策方面

在决策方面，具身智能移动操作机器人主要面临如下挑战。一是自主智能决策能力薄弱。具身智能移动操作机器人的智慧大脑需要实现环境感知与世界认知理解结果的融合匹配，依照人类价值体系生成合理、可行的决策方案，构建移动操作机器人系统可执行的指令集，完成人类执行方案与机器人系统指令集的关联映射，形成移动操作机器人解决方案。因此，移动操作机器人系统在决策时需要具备类人的价值标准、环境物体与认知物体的理解对齐能力，如将人类高级指令分解为机器人可执行指令序列的能力等，这些能力对移动操作机器人技术与具身智能发展提出了挑战。当前，具身智能决策方法主 要围绕大语言模型进行认知理解与推断，加入先验 信息引导或加入后验信息反馈矫正。由于大语言模型存在的幻觉、伪逻辑等问题，可能生成一系列机器人难以执行的指令；同时，在引入先验信息时，由于具身智能移动操作机器人的工作场景及任务通常需要即时决策，对细粒度图像与点云进行分割、特征提取的现有方法虽然效果与泛化能力很强，但计算成本与耗时均较高（如SAM[3] 5, 36），不利于即时导航与决策。近期，LLaMA[7] 、LLaVA$_1$ [23] 等通过预训练模型、有监督微调、近端优化等方式，其性能得到进一步提高，但依然存在指令通用性太强、无法具体执行的问题。

（四）运动与操作的联合规划方面

传统运动导航与机械臂运动规划相结合的方式存在灵活性差、稳定度低、空间路径次优等诸多问题。一是在复杂的室内外场景，如家庭室内环境、工厂车间环境、灾后搜救环境等，对目标的寻找、捕获和运输等任务存在很强的空间、时间条件约束，现有算法存在运动不连续、不自然、不稳定、不安全的问题。二是在动态环境中，基于局部感知进行移动导航与操作存在规划求解缓慢、效率较低等问题。虽然在静态和已知环境中的机器人导航方面已经开展了大量研究，但当机器人从静态或受控环境转变为动态环境时，多个智能体以不同的模式移动，会面临诸多挑战。三是在移动操作机器人与公共空间中其他动态机器人或人类进行交互时，对安全性提出了新的更高要求。人类可以依靠常识和经验来理解其他智能体的行为，但当前的智能移动操作机器人仅限于执行预设的交互模型完成与有限人类行为的交互，且移动操作机器人行为的有效性需要在人机系统中进行评估34。然而，由于人机混合系统的高度复杂性，难以进行理论分析和全面仿真评估；在试验评估中，人的安全性至关重要，与人相关的失败容忍度极低，因此开发更安全有效的人机交互系统评估平台至关重要。

（五）通用仿真实验平台方面

目前，移动操作机器人领域技术研发过程中缺

公众号 · 智能机器人科技

乏通用、可靠、多接口的多场景综合仿真实验平台，技术研发速度减缓，衍生出诸多实物部署中需要面临的兼容性、稳定性、准确性、泛化性问题。具身智能移动操作机器人的技术更新与迭代速度快，构建软硬件上下游系统的成本较高，但由于仿真环境多样性的限制，适用场景十分有限，亟须在仿真平台进行研发与实验验证，推进技术落地与产业化。目前，学界与业界缺乏具备通用接口、真实物理引擎、包含多场景的机器人综合仿真平台，阻碍了具身智能移动操作机器人的发展，成为制约具身智能移动操作机器人发展的一大关键节点。因此，在当前具身智能发展的浪潮下，移动操作机器人的系统研发对综合仿真实验平台的需求更加迫切。

五、基于具身智能的移动操作机器人系统发展建议

（一）紧抓具身智能移动操作机器人持续发展与产 业生态构建

我国具身智能移动操作机器人发展应着眼形成涵盖优质数据、前沿关键技术、实验测试平台、实体开发与成果转化的科技研发与落地全链条，与国内外高水平研发机构、高科技领军企业、交叉领域相关单位充分联合，有效凝聚科技协同创新的战略性资源，探索高效的有机协同机制，形成“产学研用”一体化的优势互补发展模式；以国家重大战略需求与政策为引导，推进我国基础技术、关键技术、应用部署技术的研究与突破，形成跨领域的产业化发展优势，满足国家建设、社会经济发展的战略需求，加速推进智慧社会与智慧城市的构建。

（二）注重关键技术领域的原创性突破

围绕具身智能移动操作机器人的共性关键技术，在环境感知、认知理解、智能决策、运动与操作联合规划等核心技术方面进行突破。建议行业主管部门发布具身智能共性关键技术研究计划，加强行业对新技术态势下的移动操作机器人系统的智能化、自主化、通用化、灵巧化、安全化的重视程度，切实推动具身智能发展。加速AI 技术对移动操作机器人的赋能，构建完备的软硬件交互平台，持续推进具备自动感知、自主决策、自然交互、安全执行能力的移动操作机器人系统构建。

（三）重视发展智能科学与机器人领域，关注智能 科学与技术学科建设和人才培养

推进校企联合开展关键技术研究与突破，整合创新资源，加强研究机构与上下游企业的深度合作与资源共享，围绕共性关键技术研究与工程化集成应用突破，构建具身智能移动操作机器人的创新链条与产业链条。同时，进一步完善学科领域布局，推进智能科学与技术的一级学科建设，设立智能机器人专业，增加智能机器人相关硕士研究生和博士研究生的招生名额；完善智能机器人相关学科的培养与教学方案，拓宽智能机器人学科交叉范围，与数学、物理学、生物学、计算科学等学科教育进行交叉融合。强化实验实践与实物部署在教学中的占比，提升AI技术对实体移动操作机器人的部署能力，为国家智能机器人领域发展提供充足的人才资源。

（五）鼓励建设多场景移动操作机器人系统通用验证平台

与医疗、教育、消防、交通、家庭服务、工业生产等领域中的相关单位展开深入合作，为具身智能移动操作机器人提供应用场景与示范验证平台，推进多场景智能机器人的通用性测试验证基地建设，为智能移动操作机器人发展提供充足的研究数据，为社会智能升级积累经验。借助我国与“一带一路”及其他地区的经贸合作关系，推广我国具身智能移动操作机器人的相关技术与产品，扩大领域影响力，推动形成技术与产品的全球布局。

（六）协调具身智能移动操作机器人与人类社会和谐发展

具身智能移动操作机器人技术在发展过程中因其自主性、智能性、交互性等特点，在人类社会中安全、可靠、稳定的运作依然需要在伦理、法律层面予以保障，进而确保相关技术的安全研发。在伦理层面，对于移动操作机器人的自主决策过程，技术研发需恪守白盒方式，明确机器人的完整推理与决策过程，保障人类的可靠干预，确保机器人的行为正常无风险。在法律层面，针对具身智能可能对社会带来的影响，充分考虑并予以法律保障。例如，对于移动操作机器人的自主决策与运动规划行为，需明确界定其法理责任；对于移动操作机器人

公众号 · 智能机器人科技

可能替代的行业人群，需从法律角度协调机器人的生产进程，保障人类公民的就业与生存等。配合明晰机器人行为生成的相关技术模块，便于明确责任人和责任关系，为法律执行提供合理的判罚标准与论据。

利益冲突声明

本文作者在此声明彼此之间不存在任何利益冲突或财务冲突。

Received date: November 9,2023;**Revised date:**December 15,2023

Corresponding author:Zhang Tao is a professor from the Department of Automation,Tsinghua University.His major research fields include robots and artificial intelligence.Email:taozhang@tsinghua.edu.cn

Funding project: Chinese Academy of Engineering project“Strategic Research on the New Generation of Artificial Intelligence and Industrial Clusters”（2022-PP-07）

参考文献

[1] Roa MA, Berenson D, Huang W. Mobile manipulation: Toward smart manufacturing [TC spotlight][J]. IEEE Robotics & mation Magazine, 2015, 22(4): 14-15.

[2] 刘华平，郭迪，孙富春，等. 基于形态的具身智能研究：历史回顾与前沿进展[J]. 自动化学报，2023，49(6)：1131-1154.

Liu HP, Guo D, Sun FC, et al. Morphology-based embodied intelligence: Historical retrospect and research progress [J]. Acta Automatica Sinica, 2023, 49(6): 1131-1154.

[3] 李延真，石立国，徐志根，等. 移动机器人视觉SLAM研究综述[J]. 智能计算机与应用，2022，12(7)：40-45.

Li YZ, Shi LG,Xu Z G, et al. A review of vision SLAM ods on mobile robot [J]. Intelligent Computer and Applications, 2022, 12(7): 40-45.

[4] 左晋，张皓，远子涵，等. 移动机器人SLAM发展现状综述[J]. 北京印刷学院学报，2023, 31(6):30-32.

Zuo J,ZhangH,Yuan ZH,et al.A review on the development status of mobile robot and SLAM[J].Journal of Beijing Institute of Graphic Communication,2023,31(6):30-32.

[5] 李敖. 基于SLAM的移动机器人避障研究[J].自动化应用，2023,64(3):23-26.

Li A.Research on obstacle avoidance of mobile robot based on SLAM[J].Automation Application,2023,64(3):23-26.

[6] Rubio F,Valero F,Llopis-Albert C.A review of mobile robots: Concepts,methods,theoretical framework,and applications [J]. Intemational Journal of Advanced Robotic Systems,2019,16(2): 172988141983959.

[7] Ichnowski J,Avigal Y,Kerr J,et al.Dex-NeRF:Using a neural radiance field to grasp transparent objects [EB/OL].(2021-10-27)[2023-08-30].http://arxiv.org/abs/2110.14217.pdf.

[8] Brahmbhatt S,Ham C,Kemp C C,et al.ContactDB:Analyzing and predicting grasp contact via thermal imaging [C].Long Beach:2019IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR),2019.

[9] ZeYJ,Liu YY,Shi R Z,et al.H-InDex:Visual reinforcement leaming with hand-informed representations for dexterous manipulation [EB/OL].(2023-10-02)[2023-11-09].http://arxiv.org/abs/2310.01404.pdf.

[10]Kadi HA,Terzic K.Data-driven robotic manipulation of cloth-like deformable objects:The present,challenges and future prospects [J].Sensors,2023,23(5):2389.

[11]孟繁科. 具身智能：智能进化的新阶段[J].中国工业和信息化，2023(7):6-10.

Meng F K.Embodied intelligence:A new stage of intelligence evolution [J].China Industry &Information Technology,2023(7): 6-10.

[12]谭民，王硕，机器人技术研究进展[J]. 自动化学报，2013, 39(7)：963-972.

Tan M,Wang S.Research progress on robotics [J].Acta Automatica Sinica,2013,39(7):963-972.

[13]Zhu Q,Zhang F,Huang Y,et al.An all-round AI-chemist with a scientific mind[J].National Science Review,2022,9(10):nwac190.

[14].Turing AM.Computing machinery and intelligence [M].Dordrecht:Springer Netherlands,2007:23-65.

[15]Open AI.GPT-4 [EB/OL].[2023-03-14].https://openai.com/research/gpt-4.

[16]Sun Y,Wang S,Feng S,et al.Emie 3.0:Large-scale knowledge enhanced pre-training for language understanding and generation [EB/OL].(2021-07-05)[2023-12-01].https://doi.org/10.48550/arXiv.2107.02137.

[17]Bai J,Bai S,Yang S,et al.Qwen-vl:A frontier large vision-language model with versatile abilities [EB/OL].(2023-08-24)[2023-12-01].https://doi org/10.48550/arXiv.2308.12966.

[18]Betker J,Goh G,Jing L,et al.Improving image generation with better captions [EB/OL].[2023-04-29].https://edn.openai.com/papers/dall-e-3.pdf.

[19]Team G,Anil R,Borgeaud S,et al.Gemini:A family of highly capable multimodal models [EB/OL].(2023-12-19)[2023-12-20].https://doi org/10.48550/arXiv.2312.11805.

[20]Luo H,Sun Q,Xu C,et al.Wizardmath:Empowering mathematical reasoning for large language models via reinforced evol-instruct [EB/OL].(2023-08-18)[2023-09-20].https://doi.org/10.48550/arXiv.2308.09583.

[21]Driess D,XiaF,Sajjadi MSM,et al.PaLM-E:Anembodied multimodal language model [EB/OL].(2023-03-06)[2023-11-10].http://arxiv.org/abs/2303.03378.pdf.

[22]Liu HT,LiCY,WuQY,et al.Visual instruction tuning [EB/OL].(2023-04-17)[2023-11-09].http://arxiv.org/abs/2304.08485.pdf.

[23]Wu HT,Jing Y,Cheang CH,et al.Unleashing large-scale video generative pre-training for visual robot manipulation [EB/OL].(2023-12-20)[2023-12-21].https://doi.org/10.48550/arXiv.2312.13139.

[24]Fu ZP,Zhao TZ,Finn C.Mobile ALOHA:Learning bimanual mobile manipulation with low-cost whole-body teleoperation [EB/OL].(2024-01-04)[2024-01-06].https://doiorg/10.48550/arXiv.2401.02117.

[25]Gu J,Kirmani S,Wohlhart P,et al.RT-Trajectory:Robotic task generalization via hindsight trajectory sketches [EB/OL].(2023-11-03)[2023-12-21].https://doi.org/10.48550/arXiv.2311.01977.

公众号·智能机器人科技

[26]多可移动协作机器人[EB/OL].[2023-10-18].https://www.ducorobots.cn/index/goods/prodetail/cid/31.html.
DUCO mobile cobot [EB/OL].[2023-10-18].https://www.ducorobots.cn/index/goods/prodetail/cid/3L html

[27]迦智科技. 复合移动作业机器人[EB/OL].[2023-10-18].https:lwww.iplusbot.cn/product/184-271.
IPLUS Mobot.Composite mobile operation robot [EB/OL].[2023-10-18].https://www iplusbot.cn/product/184-271.

[28] KUKA.KUKA 移动机器人[EB/OL].[2023-10-18].https:I/www.kuka.com/zh-cn/%E4%BA%A7%E5%93%81,-a-,%E6%88%90%E6%9E%9C%E6%9C%BA%E5%8A%A8%E6%80%A7/%E7%A7%BB%E5%8A%A8%E6%9C%BA%E5%99%A8%E4%BA%BA.
KUKA.KUKA autonomous mobile robots [EB/OL].[2023-10-18].https://www.kuka com/zh-cn/%E4%BA%A7%E5%93%81,-a-,%E6%88%90%E6%9E%9C/%E6%9C%BA%E5%8A%A8%E6%80%A7/%E7%A7%BB%E5%8A%A8%E6%9C%BA%E5%99%A8%E4%BA%BA.

[29]朱悦. 谷歌DeepMind发布机器人大模型RT-2,提高泛化与涌现能 力[EB/OL].(2023-08-02)[2023-12-10].https://baijiahao.baidu.com/s?id=1773102794187718464&wfi=spider&for=pc.
Zhu Y.Google DeepMind released the robot model RT-2 to improve the generalization and emergence capabilities [EB/OL].(2023-08-02)[2023-12-10].https:/baijiahao.baidu.com/s?id=1773102794187718464&wfi=spider&for=pc.

[30]1X.Data collection for embodied learning [EB/OL].[2023-10-11].https://www.Ix.tech/androids/eve.

[31]Boston Dynamics.Meta:Advanced AI research [EB/OL].[2023-02-01].https//bostondynamics.com/case-studies/advanced-ai-adding-capabilities-to-spot-through-research/

[32]罗欣，丁晓军. 地面移动作业机器人运动规划与控制研究综述[J]. 哈尔滨工业大学学报，2021, 53(1):1-15.
Luo X,Ding X J.Research and prospective on motion planning and control of ground mobile manipulators [J].Journal of Harbin Institute of Technology,2021,53(1):1-15.

[33]Edmonds M,Gao F,LiuHX,et al.A tale of two explanations:Enhancing human trust by explaining robot behavior [J].Science Robotics,2019,4(37):eaay4663.

[34]Mavrogiannis C,Baldini F,Wang A,et al Core challenges of social robot navigation:A survey [J].ACM Transactions on Human-Robot Interaction,12(3):36.

[35]Kirillov A,Mintun E,Ravi N,et al.Segment anything [EB/OL].(2023-04-05)[2023-11-09].http://arxiv.org/abs/2304.02643.pdf.

[36]Zheng Y,Wang GM,LiuJM,et al.Spherical frustum sparse convolution network for LiDAR point cloud semantic segmentation [EB/OL].(2023-11-29)[2023-11-30].http://arxiv.org/abs/2311.17491.pdf.

[37]TouvronH,Lavril T,Izacard G,et al.LLaMA:Open and efficient foundation language models [EB/OL].(2023-02-27)[2023-11-09].http://arxiv.org/abs/2302.13971.pdf.

[38]徐继敏. 生成式AI治理原则与法律策略[J].理论与改革，2023(5):72-83.
Xu JM.Governance principles and legal strategies of generative artificial intelligence[J].Theory and Reform,2023(5):72-83.

公众号 · 智能机器人科技