

DOI:10.11784/tdxbz202312024

基于视觉定位与虚拟现实技术的机器人遥操作研究

林子骞¹, 肖聚亮¹, 赵 炜¹, 刘海涛¹, 龙晖午¹, 牛一凡²

(1. 天津大学机构理论与装备设计教育部重点实验室, 天津 300350;

2. 北京新风航天装备有限公司, 北京 100854)

摘要: 机器人遥操作, 即人类操作者通过远程控制设备操纵机器人完成任务, 实现了机器人的执行能力和人类操作者在非结构化环境中的决策能力的融合, 已被广泛应用于各种工作要求复杂且对人类有潜在危险的领域。高效遥操作的关键因素之一在于能否为操作者提供良好的临场感, 临场感取决于操作者对远程环境的身临其境的感知(视觉、触觉等感官反馈)以及对远程机器人精准、直观的控制。本文提出了一种基于视觉定位与虚拟现实技术的机器人遥操作方法, 重建并传输远程场景的静态点云模型, 利用 Gen6D 位姿估计算法根据 RGB 数据自动检测远程场景中的物体, 然后将检测到的物体位姿信息传输到 Unity 虚拟现实场景中, 定位虚拟物体, 从而取代物体的点云模型。充分发挥虚拟现实的固有优势, 沉浸式可视化重建后的远程场景, 利用 Unity 物理引擎驱动虚拟物体, 提高操作者的临场感。其次, 提出一种人机解耦的信息物理(cyber-physical)映射模型, 将操作者与远程机器人的控制回路解耦, 消除了时延对遥操作的影响。最后, 提出一种基于虚拟现实空间的自适应位置增量映射算法, 根据虚拟机器人末端抓手与目标物体的距离改变位姿映射的比例, 提高了遥操作的灵活性, 并赋予虚拟现实遥操作避障的功能, 保障了虚拟现实遥操作的安全性。

关键词: 机器人; 遥操作; 虚拟现实; 场景重建; 位姿映射

中图分类号: TP242 **文献标志码:** A **文章编号:** 0493-2137(2025)01-0036-11

Research on Robot Teleoperation Systems Based on Visual Positioning and Virtual Reality Technology

Lin Zi'ao¹, Xiao Juliang¹, Zhao Wei¹, Liu Haitao¹, Long Huiwu¹, Niu Yifan²

(1. Key Laboratory of Mechanism Theory and Equipment Design of Ministry of Education,
Tianjin University, Tianjin 300350, China;

2. Beijing Xinfeng Aerospace Equipment Co., Ltd., Beijing 100854, China)

Abstract: Robot teleoperation, where a human operator remotely controls a robot to perform tasks, combines the decision-making capabilities of humans with the execution capabilities of robots in unstructured environments. This approach has been widely used in various fields that involve complex tasks and potential hazards for humans. Crucial for efficient teleoperation is providing the operator with a strong sense of presence, which depends on the operator's immersive perception of the remote environment, including visual, tactile, and other sensory feedback, as well as precise and intuitive control of the robot. In this paper, an innovative robot teleoperation method that leverages visual localization and virtual reality technology is proposed. The proposed method begins by reconstructing and transmitting a static point-cloud model of the remote scene. Using the Gen6D pose estimation algorithm, objects in the scene are automatically detected based on RGB data. The pose information of these detected objects is then transmitted to a Unity virtual reality scene, where the point-cloud models of the objects are replaced with virtual equivalents. We fully leverage the inherent advantages of virtual reality to visualize the reconstructed remote scene immersively. The Unity

收稿日期: 2023-12-18; 修回日期: 2024-01-18.

作者简介: 林子骞(1999—), 男, 博士研究生, 1031958775@qq.com.

通信作者: 肖聚亮, tianjinxl@163.com.

基金项目: 国家自然科学基金资助项目(52175025, 91948301).

Supported by the National Natural Science Foundation of China (No. 52175025, No. 91948301).

physics engine drives the virtual objects, improving the operator's sense of presence. Moreover, we propose a cyber-physical mapping model with human-machine decoupling, which separates the control loops of the operator and the robot, effectively mitigating the impact of time delays on remote operations. We also propose an adaptive position incremental mapping algorithm within the virtual reality space. This algorithm adjusts the scale of pose mapping according to the distance between the virtual robot's end gripper and the target object, thereby improving teleoperation flexibility. This also incorporates obstacle avoidance features, ensuring the safety of teleoperation within the virtual reality environment.

Keywords: robot; teleoperation; virtual reality; scene reconstruction; pose mapping

机器人遥操作,即人类操作者通过远程控制设备操纵机器人完成任务,实现了机器人的执行能力和人类操作者在非结构化环境中的决策能力的融合^[1]。目前,遥操作技术已被广泛应用于各种工作要求复杂且对人类有潜在危险的领域,例如排雷排爆、搜索救援、高空作业、水下作业和太空探索等。

在实际应用中实现直观、高效遥操作的关键因素之一在于能否为操作者提供良好的临场感^[2]。临场感,即让操作者充分感知远程机器人环境,从而在需要迅速反应的情况下做出更加准确和高效的决策。临场感又可分为视觉与力触觉临场感^[3],其中视觉临场感在遥操作中的重要性已得到广泛的认识并得到较为深入的研究,因为视觉反馈是遥操作中最重要感官反馈^[4]。视觉临场感主要由视觉反馈的形式和传输的延迟决定。传统遥操作中的视觉反馈主要依赖于来自远程环境的多角度 2D 视频,其优点是信息来源简单直接、容易获取^[5];其局限性在于操作者在切换不同角度的视频时,需要很强的坐标变换能力和空间转换能力,负荷较大、效率较低。此外,由于深度信息的缺失,往往会引起人对深度距离的错误感知^[6]。同时,基于 2D 视频的视觉反馈形式无法避免信息传输、图形计算过程中造成的时延。以上两点导致无法提供给操作者良好的临场感,极大地降低了遥操作的效率,同时会给遥操作的安全性带来影响^[7]。

随着虚拟现实(virtual reality, VR)技术的发展^[8],其已被证实是提高遥操作工作效率、提升临场感的关键技术^[9]。近年来,国内外已提出了各种不同的基于虚拟现实视觉反馈的机器人遥操作系统。目前大多数虚拟现实遥操作系统通常选用廉价的 RGB-D 相机^[10]捕获 2D 视频^[11]和点云^[12]实现远程场景重建。例如, Naceri 等^[13]开发的 Vicarios 虚拟现实遥操作系统在虚拟现实利用高质量的实时 2D 视频流和实时点云的可视化组合对远程环境以及远程机器人进行一对一的立体渲染,改善了操作人员对远程场景的深度感知,提升了操作者的临场感。Xu 等^[14]提出了一种基于虚拟现实的人机共享控制遥操作系统,同样是

采用点云和视频反馈相结合的方式,并通过基于路径点的控制方法提高了遥操作的流畅性。尽管上述基于点云与 2D 视频的虚拟现实视觉反馈方式对遥操作的直观性和效率有明显提升,但还是存在改进的空间。点云模型的数据实时传输延迟较大^[15],目前大多数基于点云重建的虚拟现实遥操作系统无法在虚拟现实设备显示中达到令人满意的刷新率^[16]。操作者长时间面对大时延、低帧率的虚拟现实视觉反馈时,会出现恶心和头晕的问题^[17]。针对点云模型量大、传输效率低的问题,Zhou 等^[18]提出了基于深度学习重建远程场景的机器人遥操作系统(teleoperation based on automatic reconstruction of remote scenes, TOARS),利用深度学习算法根据点云模型自动检测远程场景中的物体及其物理特性,然后将处理后的信息传输到 Unity 引擎中,在 Unity 中使用虚拟物体取代原始的点云模型。TOARS 显著提高了机器人遥操作中人类操作者的态势感知能力,减轻了点云传输的压力。但 TOARS 遥操作系统采用 PointNet 深度学习算法对点云进行分类,该算法分类点云需提前制作大量的待识别物体的数据集,泛化性较差,很难应对非结构化环境;并且 TOARS 系统采用传统的将操作者与机器人的控制回路耦合的直接映射模型(direct model)^[19]来映射人类操作者的输入与机器人的运动,直接将操作者手持控制器的位姿映射到远程机器人上,再从远端将机器人状态信息传回虚拟现实环境中来更新虚拟机器人的位姿,这个过程中数据传输造成的延迟会导致远程机器人状态信息的视觉反馈无法与手持控制器的动作匹配,导致使用者疲劳、恶心,降低了操作者遥操作的效率。

针对上述问题,本文提出了一种基于视觉定位与虚拟现实技术的机器人遥操作方法。结合点云与泛化性较强的 Gen6D 位姿识别算法实现对远程场景的重建,提高了对非结构化环境的适应能力。提出一种人机解耦的信息物理映射模型,将操作者与远程机器人的控制回路解耦,消除了时延影响。提出一种基于虚拟现实空间的自适应位置增量映射算法,根据虚拟空

间中机器人末端与目标物体的距离改变位姿映射的比例,提高了操作效率与灵活性,并赋予虚拟现实遥操作系统避障的功能,提高了虚拟现实遥操作的安全性.

1 系统框架

图 1 简洁直观地描述了基于虚拟现实技术的遥操作中直接映射模型与本文提出的人机解耦的信息物理映射模型之间的区别. 每个对象的位姿在指定的参考系中由一个变换矩阵表示,用上标来表示每个对

象的参考系. 在虚拟现实机器人遥操作系统中,需要有选择地在操作者参考系 U 、虚拟现实参考系 V 和远程机器人参考系 R 之间进行映射,其中包括每个对象的位姿以及其他的状态信息,映射方法将直接影响遥操作系统的精度以及效率.

在图 1(a)所示的直接映射模型中操作者直接对远程机器人进行控制,操作者接收到的实时视觉反馈与远程机器人在虚拟场景中的状态(S^V)相耦合,由于机器人运动必然带有一定的滞后性且伴随信息流传递与画面渲染都需要消耗一定的时间,因此在目前视觉反馈人机耦合的状态下必然面临着因遥操作存在延迟而使得操作者存在晕眩反应的问题,影响了操作的效率.

如图 1(b)所示,本文基于人机解耦的信息物理映射模型的虚拟现实遥操作系统为操作者和机器人创建共享的虚拟场景,基于远程机器人的先验信息在虚拟场景中创建两个虚拟机器人替身,其中虚拟机器人 1(S_1^V)能灵活地跟随操作者手持控制器的位姿,实时地可视化操作者下达的命令,实现对远程机器人的轨迹预测;虚拟机器人 2(S_2^V)则负责将远程机器人传回的位姿信息实时地反馈给操作者. 在这个过程中将向操作者提供的感官反馈与机器人的通信分离开来,实现了操作者与远程机器人控制回路的解耦,时延转化为 2 个虚拟机器人的先后运动,消除了时延影响.

本文遥操作系统的具体框架如图 2 所示.

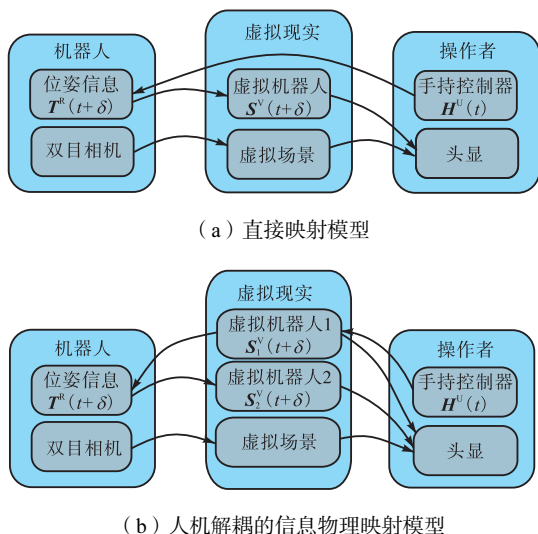


图 1 虚拟现实遥操作映射模型

Fig.1 Virtual reality-based teleoperation mapping models

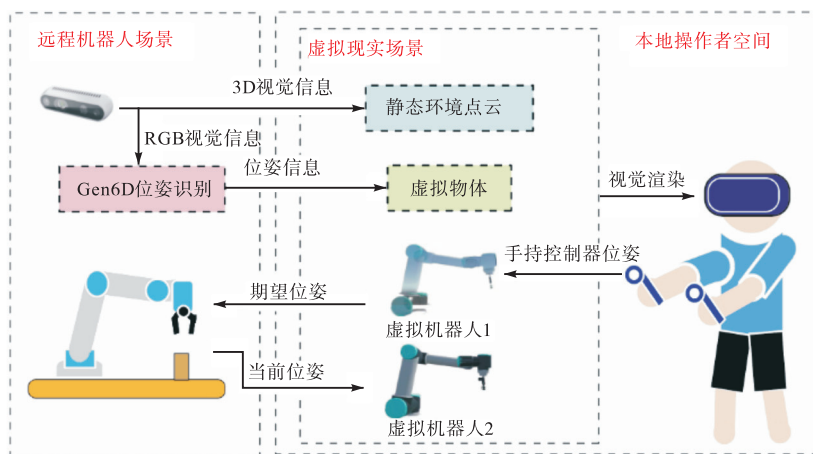


图 2 虚拟现实机器人遥操作系统框架

Fig.2 Framework for virtual reality-based robot teleoperation system

遥操作系统主要分为远程机器人场景与本地操作者空间,其中远程机器人场景由 RGB-D 相机和机器人构成,本地操作者空间由手持控制器并佩戴虚拟现实设备的操作者和通过 Unity 引擎创建的虚拟现

实场景构成. 其中虚拟现实场景负责远程机器人场景的重建和主从两端的信息传输,是框架的关键部分.

虚拟现实场景包括: RGB-D 相机捕获的远程机器人场景静态环境点云;替代原始点云模型的虚拟物

体；基于远程机器人的先验信息创建的虚拟机器人替身。操作者佩戴虚拟现实设备沉浸在由 Unity 构建的虚拟现实场景中，通过使用手持控制器来控制虚拟机器人 1，其位姿通过 TCP 通讯发送给远程机器人，在真实的机器人上再现相同的运动。同时远程机器人的位姿传回虚拟现实场景，通过虚拟机器人 2 反馈给操作者。

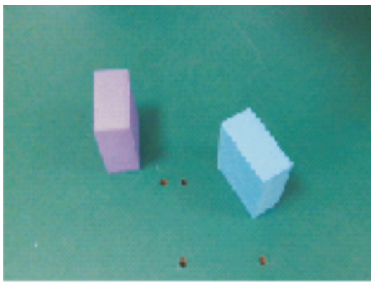
2 基于点云与 Gen6D 位姿估计的场景重建方法

2.1 远程环境点云的捕获以及可视化

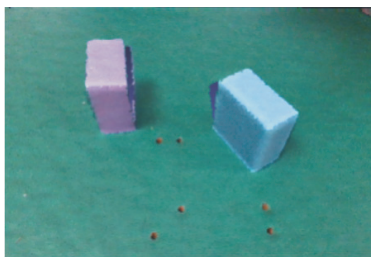
捕获远程机器人空间的三维点云是虚拟现实场景重建的第一步，为了减轻计算机和网络传输的负担，应尽量减少点云传输的数据量，如图 3 所示。采用 Realsense D435i 深度相机捕获较低分辨率 (640 像素 × 480 像素) 的远程环境三维点云，通过 TCP 通讯发送到操作者空间的 Unity 虚拟现实场景，在 Unity



(a) RGB 图像



(b) 深度图像



(c) 三维点云

图 3 远程环境点云捕获

Fig.3 Remote environment point-cloud capture

中使用这些点云数据渲染网格，创建点云表示，实现点云的可视化。

为了在 Unity 虚拟现实场景中精确地定位点云模型位置，以使操作者看到的虚拟场景与现实世界中的场景相匹配，需要得到机器人坐标系与 Realsense 相机坐标系之间的变换矩阵，即进行深度相机的标定。深度相机标定旨在通过确定相机的内部参数和外部位姿，确保深度相机能够准确捕捉场景的深度信息，为后续的计算和应用提供可靠的数据基础。Realsense 深度相机的内参矩阵在制造时已经进行了标定，并通过其 SDK 提供调用接口。在使用过程中，只需要对外参矩阵 N 进行标定即可。

本文使用如图 4 所示的“眼在手外”的外参标定方法找到机器人和 Realsense 之间的相对姿势转换。将标定板与机器人末端固连，则相机坐标系到机器人基坐标系的转化矩阵 N 可以表示为

$$N = T_{Robot}^{Board} \cdot T_{Board}^{Realsense} \quad (1)$$

式中： T_{Robot}^{Board} 为已知的从标定板到机器人基坐标系的矩阵； $T_{Board}^{Realsense}$ 为相机坐标系到标定板的变换矩阵，可通过识别程序标定得到。

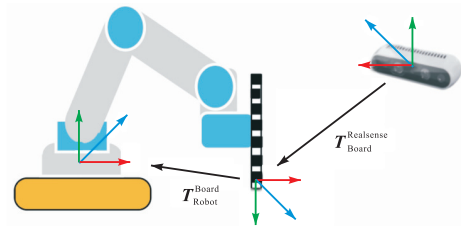


图 4 相机坐标系到机器人基坐标系的坐标转换

Fig.4 Converting coordinates from the camera coordinate system to the robot base coordinate system

完成以上步骤后，操作者可以在虚拟现实头显中看到远程场景的点云，虚拟机器人与环境的点云应该与远程环境完全匹配和重叠。

2.2 基于 Gen6D 位姿估计的虚拟物体替换

2.2.1 替换策略

如图 5(a) 所示，传统的场景重建方法必须每个时刻重建并传输远程场景点云，这个过程中造成的时延会对遥操作系统的稳定性造成影响。相比之下，如图 5(b) 所示的基于 Gen6D 位姿估计的场景重建方法仅在 t_0 时刻重建并传输远程场景点云，同时识别场景中物体位姿，通过 Unity 物理引擎驱动虚拟物体位姿，并用虚拟物体替代点云，充分利用 Unity 物理引擎的优势，减少了点云数据实时传输的负担。

图 6 展示了远程场景重建的过程。首先，通过 Realsense 相机采集远程场景的 RGB 图像与深度图

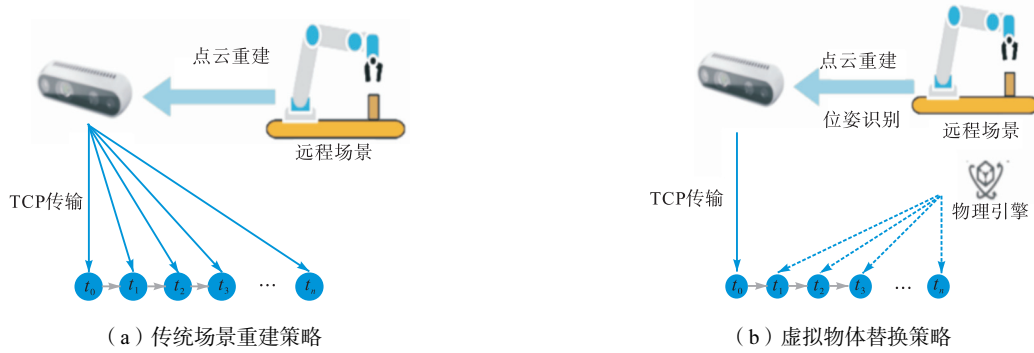


图 5 场景重建策略比较

Fig.5 Comparison between scene reconstruction strategies

像合成点云,对于远程场景中的静止对象,例如桌面、墙壁、地板等,直接使用捕获到的点云数据而不进行替换,保留环境的点云信息是为了增强操作者对远程环境的理解能力.随后通过 Gen6D 位姿估计算法根据 RGB 图像识别物体位姿,将识别到的位姿信

息通过 TCP 发送到 Unity 中.为了在 Unity 虚拟现实场景中点云数据替换为虚拟对象,基于收到的物体信息(名称、位姿)在模型库中进行搜索,随后生成一个 GameObject 类的虚拟物体并通过 Instantiate(GameObject, Vector3, Quaternion)方法调整其姿态.

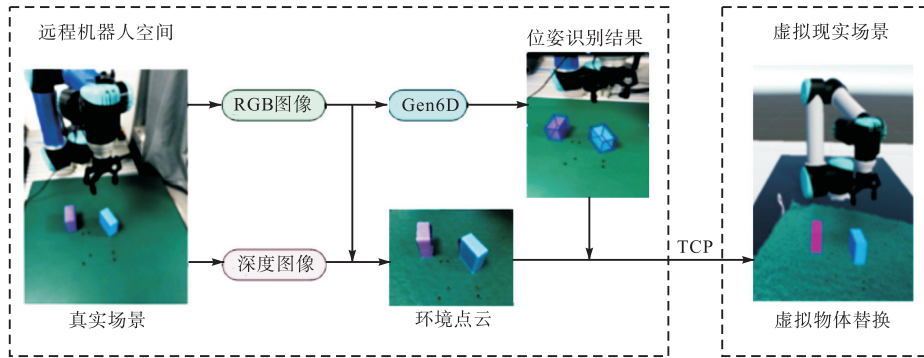


图 6 远程场景重建过程

Fig.6 Remote scene reconstruction process

2.2.2 Gen6D 位姿估计算法

Gen6D 位姿估计算法不依赖于高质量的 3D 模型或者额外的深度图和掩膜^[20],同时泛化性强,不需要对新的物体制作训练数据集进行训练,就能预测该物体在图片中的位姿.

如图 7 所示,Gen6D 位姿估计算法基于一种新的图像匹配的框架,由检测器、视点选择器和姿态精细器组成,以由粗到精的方式估计目标姿态. Gen6D 基于已知相机位姿的物体的参考图像($N_0 - N_r$),预测查询图像中物体的位姿,即将物体坐标系位姿 x_{obj} 转换为相机坐标系位姿 $x_{cam} = R x_{obj} + m$,其中 R 为旋转矩阵, m 为平移向量.其中目标检测器裁剪查询图像中的目标区域并估计目标的初始平移,视点选择器通过选择最相似的参考图像并估计平面内旋转来找到初始旋转,姿态精细器基于初始的平移,旋转迭代地估计出一个更精确的姿态.

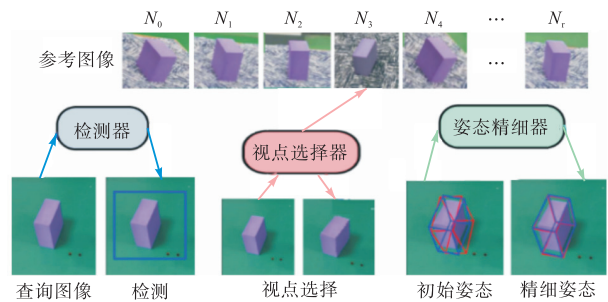


图 7 Gen6D位姿估计算法框架

Fig.7 Framework of Gen6D pose estimation algorithm

未知物体进行姿态估计时,只需首先绕着未知物体拍摄一段视频作为参考序列,随后使用 COLMAP 软件恢复出每一张参考图像的位姿,同时重建未知物体的稀疏点云以定义物体的大小,便可在短时间内实现对未知物体的姿态识别,而不需大量的时间制作数据集. Gen6D 算法强大的泛化能力满足虚拟现实遥操作系统的应⽤要求,提高了场景重建系统的工作效率.

在遥操作工作中,非结构化的工作环境导致操作者很可能面对未知的物体,在使用 Gen6D 算法对

3 虚拟现实中的远程机器人遥操作

在虚拟现实场景中根据远程机器人的先验信息创建了虚拟机器人1和虚拟机器人2两个替身,将操作者对远程机器人的控制解耦为如图8所示的3次映射 $M_{H^U S_1^V}$ 、 $M_{S_1^V T^R}$ 、 $M_{T^R S_2^V}$,其中 $M_{H^U S_1^V}$ 将操作者空间映射到虚拟现实空间, $M_{S_1^V T^R}$ 将虚拟现实空间映射到远程机器人空间, $M_{T^R S_2^V}$ 将远程机器人空间反向映射回虚拟现实空间.

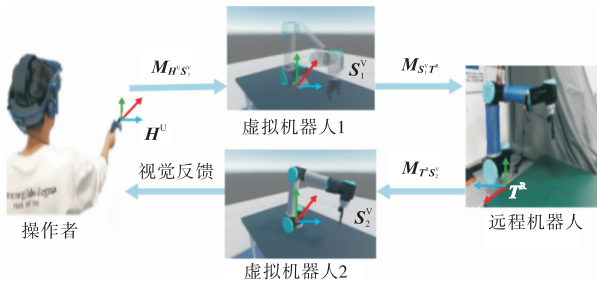


图8 远程机器人控制中的3次映射

Fig.8 Three mapping methods for remote robot control

3.1 操作者空间到虚拟现实空间的映射

上述3次映射中,如何实现操作者空间到虚拟现实空间的映射 $M_{H^U S_1^V}$,即如何将操作者手持控制器的末端位姿映射到虚拟机器人1的末端,是进行直观有效的远程机器人遥操作的关键.在面向较为复杂的任务时,使用传统的定比例映射法会导致遥操作时出现工作效率低、灵活性差等问题.如图9所示,以积木抓取任务为例,在操纵机器人从远处接近目标积木的阶段时,可使用较大的比例系数进行快速移动.接近积木后,需要精确地调整末端抓手的位姿以成功抓起积木,故需要较小的比例系数进行精细的操作.因此本文提出基于虚拟现实空间的自适应位置增量映射法完成操作者空间到虚拟现实空间的映射.

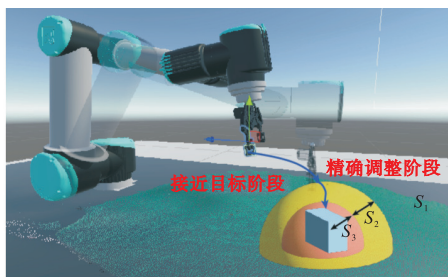


图9 虚拟现实环境中积木抓取任务操作流程

Fig.9 Procedure of grasping building blocks in a virtual reality environment

3.1.1 位置增量映射法

本文选用位置增量式映射法作的映射 $M_{H^U S_1^V}$ 基础.操作者手持控制器 t 帧的位姿 $H^U(t)$ 与虚拟机器人1在 t 帧的末端位姿 $S_1^V(t)$ 可分别表示为

$$H^U(t) = [x_h, y_h, z_h, a, b, c] \quad (2)$$

$$S_1^V(t) = [x_s, y_s, z_s, \alpha, \beta, \gamma] \quad (3)$$

式中: x_h, y_h, z_h, a, b, c 表示手持控制器在 Unity 坐标系下的位置以及欧拉角表示的姿态,可通过 Unity 内的 Vive Input Utility 插件直接得到; $x_s, y_s, z_s, \alpha, \beta, \gamma$ 表示虚拟机器人1末端在其基坐标系下的位置以及欧拉角表示的姿态,通过在 Unity 中编写的机器人正运动学计算得到.如图10所示,当操作者持续按住 HTC Vive 控制器的扳机(trigger)时,Unity 系统将记录下手持控制器 t 帧的位姿,随后操作者移动手持控制器改变其位姿时, $t+1$ 帧虚拟机器人1末端位姿可以表示为

$$S_1^V(t+1) = k[H^U(t+1) - H^U(t)] + S_1^V(t) \quad (4)$$

式中 k 为手持控制器位姿与虚拟机器人1末端位姿之间的映射比例.随后根据在 Unity 中编写的机器人逆运动学程序将 $S_1^V(t+1)$ 转化为各关节角度,实现虚拟机器人1的末端成比例地跟随控制器进行运动.当操作者停止并松开扳机时,系统就会忘记上次记录的位置,需要重新按下按钮记录另一个位置,以便进行下一次位姿映射.

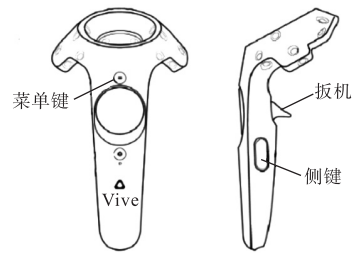


图10 HTC Vive手柄控制器按键布局

Fig.10 Button layout of HTC Vive handle controller

3.1.2 基于虚拟现实空间的自适应映射比例调节

当映射比例系数 k 较大时,虚拟机器人可快速接近任务目标. k 较小时,更利于操作者操控虚拟机器人完成精细化操作.比例系数应根据执行任务不同阶段对精细化程度的需求进行切换,以提高遥操作系统的灵活性.传统的比例系数调节方法是由操作者主动切换,操作者在主端通过按钮直接控制比例系数的切换,该方法简单易实现,但是对于多个比例系数之间切换则需要设置更多的按钮,设置的按钮数量过多会影响操作者的体验,干扰操作者对任务本身的注意

力. 为了在保证遥操作体验的同时实现比例系数流畅、合理的切换, 并充分发挥 Unity 虚拟现实空间的优势, 本文提出了一种基于虚拟现实空间的自适应映射比例调节方法. 考虑到在 Unity 物理引擎中, 虚拟机器人末端抓手与抓取目标的距离可以实时读取, 所以选择基于虚拟现实场景中的距离实现映射比例的自然调节. 如图 9 所示, 在虚拟现实环境中以虚拟积木为中心、半径分别为 r_0 、 r_1 (半径具体数值由抓取目标的大小而定) 的两个半球体划分出了 3 个工作空间 S_1 、 S_2 、 S_3 . 这 3 个工作空间的映射比例系数分别为 k_1 、 k_2 、 k_3 . 其中, k_1 、 k_2 均被设为定值 ($k_1=0.8$, $k_2=0.3$), 因为在同一任务阶段中如果比例系数忽大忽小, 对于操作者来说操作体验不佳.

在实际应用过程中, 从机器人接近目标物和环境交互的过程不仅需要考虑遥操作控制的运动精度, 安全性也是重要的考虑因素. 当虚拟机器人末端抓手进入 S_3 工作空间时, 意味着抓手与目标物体过于接近, 存在碰撞的风险, 于是 S_3 的“斥力”将会驱使 k_3 不断减小直至为 0, 防止与目标物体发生碰撞, 比例系数 k_3 的函数可以表示为

$$k_3 = k_2 e^{-\frac{r_0}{r}} \quad 0 < r \leq r_0 \quad (5)$$

式中: r_0 为工作空间 S_3 所在半球的半径; r 为虚拟机器人抓手到虚拟积木的距离. 上述基于虚拟现实空间的自适应位置增量映射法的实现过程如图 11 所示. 同时操作者可通过如图 10 所示的菜单按钮实现抓手在打开和关闭状态之间的切换. 当操作者在虚拟现实场景中按下按钮抓住虚拟积木时, 虚拟积木将变为虚拟机器人抓手的子物体, 虚拟积木对抓手的“斥力”也会随之消失.

3.2 虚拟现实空间到远程机器人空间的映射

虚拟机器人 1 与远程机器人之间的映射 $M_{S^V T^R}$ 采用了基于路径点的映射方法, 其类似于使用电脑鼠标的点击. 在用户按下 HTC Vive 控制器扳机通过映射 M_{V^R} 将虚拟机器人 1 移动到自己期望的位置后, 相当于使用手持控制器这个“鼠标”为远程机器人制定了期望的路径点. 如图 10 所示, 随后按下 HTC Vive 控制器侧键, 将虚拟机器人 1 状态信息(6 个关节角度和抓手状态)通过 TCP 通讯发送给远程机器人, 驱动其动作, 实现虚拟机器人 1 与远程机器人之间的映射.

3.3 远程机器人空间到虚拟现实空间的反向映射

远程机器人与虚拟机器人 2 的位姿映射 $M_{T^R S^V}$ 指的是远程机器人将通过 TCP 通讯实时地将状态信

息传回 Unity 虚拟环境, 虚拟机器人 2 将根据传回的数据更新自身姿态, 这使得操作者同时能够直观地获取远程机器人位姿的视觉反馈.

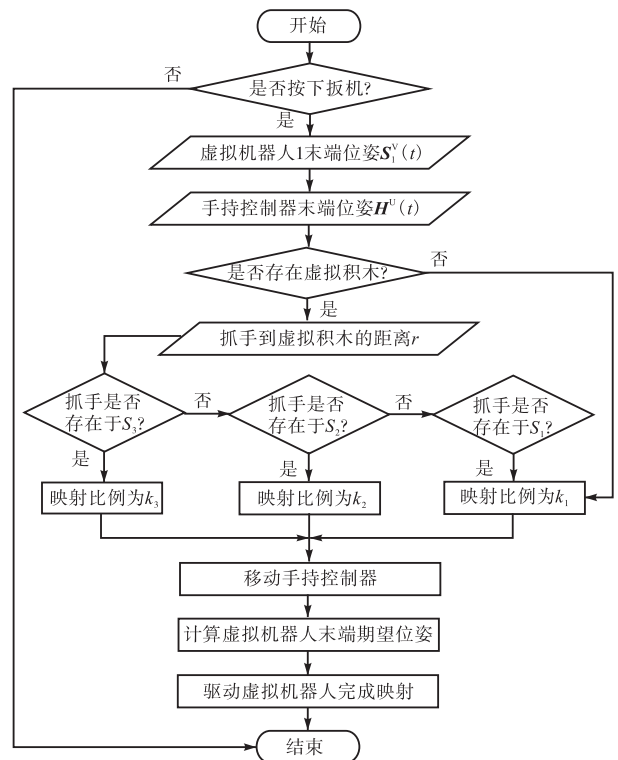


图 11 基于虚拟现实空间的变比例位置增量映射法流程
Fig.11 Flow chart of variable-scale location incremental mapping method based on virtual reality space

4 实验验证

4.1 系统搭建

通过积木堆叠实验验证本文提出的虚拟现实遥操作系统的有效性, 首先需完成实验平台的搭建. 虚拟现实机器人遥操作系统的构建包括本地操作者空间、远程机器人场景 2 个部分, 其中主端操作者空间如图 12 所示.

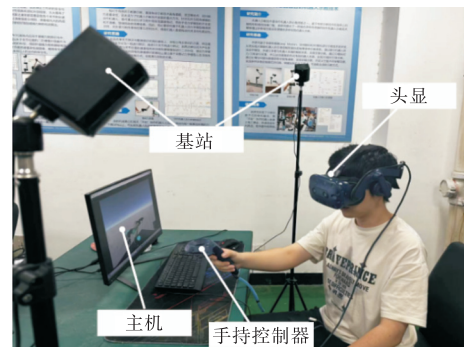


图 12 操作者空间组成
Fig.12 Composition of local operator space

本地操作者空间由以下部分组成。

(1) 虚拟现实设备 HTC Vive Pro, 包括一个头戴显示器和一对手持无线控制器。头戴显示器分辨率为 2880 像素 × 1600 像素, 拥有 110° 视野, 90 Hz 刷新率。其独有的灯塔跟踪定位系统, 可以实现精确的头部运动跟踪。两个控制器都可以相对于头戴显示器进行 6 自由度跟踪。头戴显示器通过 USB-C 连接到本地计算机, 同时手持无线控制器通过蓝牙与头戴显示器连接。

(2) 一台运行 Unity 物理引擎的计算机。Unity 内的虚拟场景数据, 包括远程机器人状态、环境点云以及位姿估计的结果, 均使用 TCP 通讯从远程机器人场景传输到本地。远程机器人的状态数据, 包括其 URDF 模型、关节角度和抓手状态用于在虚拟环境中构建虚拟机器人。远程环境点云在 Unity 中重建远程环境的固定对象。随后通过位姿估计的结果在 Unity 中用虚拟模型代替其点云。除此之外, 在遥操作的过程中向远程机器人发送的命令也在 Unity 中通过 TCP 通讯进行传输。

从端远程机器人场景如图 13 所示, 由以下部分组成。

(1) 遥操作系统使用的是实验室自主研发的协作机器人 R-Sv5, 机器人运动控制器为 Delta tau 公司的 PowerPMAC。PowerPMAC 可以实现多种通信方式, 包括 EtherCAT、串口通信和 TCP 通信等。通过 EtherCAT 接口连接 PowerPMAC 与电机驱动器, 实现机器人各关节的同步运动, 通过 TCP 通信实现 PowerPMAC 与 Unity 虚拟空间的数据传递。PowerPMAC 具有计算速度快、精度高、可同步序列运动程序、可同步收集数据等优势。同时机械臂末端还安装了一个低成本抓手, 连接在 PowerPMAC 控制器上通过串口实现通信。

(2) Realsense D435i 双目摄像机, 该相机可以以 30 帧/s 的速度捕获 1920 像素 × 1080 像素分辨率的彩色图像数据, 并以 90 帧/s 的速度捕获 512 像素 × 424 像素分辨率的深度图像数据。采用 Realsense 自带的 SDK 对其进行融合生成点云数据。并通过 TCP 通讯将 Realsense 摄像头捕获的点云数据发送到 Unity 虚拟场景。

(3) 一台运行 Gen6D 位姿估计算法, 与点云传输的远端计算机, 搭载了 4090GPU, Intel i9-13900K CPU。该计算机负责通过借助 Realsense D435i 双目相机捕获到的 RGB 图像进行 Gen6D 位姿估计, 将识别到的结果发送到本地操作者空间, 同时也负责完成

点云模型的合成和发送。

在远程工作场景与操作者空间之间通过局域网连接。基于 TCP 通讯协议建立客户端(操作空间)与服务器端(远程环境)的通讯。局域网的平均下载速度为 51.67 Mb/s, 上传速度为 44.52 Mb/s, 平均网络延迟为 3 ms。

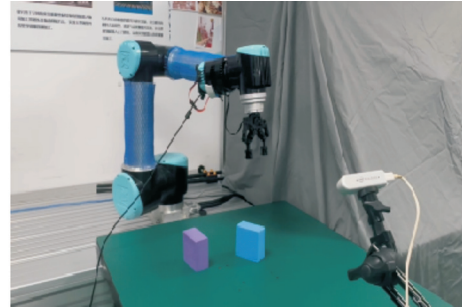


图 13 远程机器人场景组成

Fig.13 Composition of remote robot scene

4.2 实验设计

本节将进行如图 14 所示的遥操作积木堆叠实验, 评估虚拟现实遥操作系统对操作者在遥操作任务中的操作效率和准确性的影响。在这个实验中操作者的任务是移动机器人抓取紫色积木, 随后移动机器人到蓝色积木上方并释放抓手实现积木堆叠。在实验开始前, 通过基于点云与 Gen6D 位姿估计的场景重建方法在 Unity 中搭建虚拟场景, 具体重建流程不再赘述。设置工作空间 S_2 的阈值 $r_1 = 10$ cm, 工作空间 S_3 的阈值 $r_2 = 8$ cm。操作者将在 3 种不同条件下进行积木堆叠实验。

(1) 采用传统的人机耦合的直接映射模型, 同时采用传统的固定比例的位姿增量映射方法将操作者手持控制器位姿直接映射到远程机器人。

(2) 采用人机解耦的信息物理映射模型, 但是采用传统的固定比例的位置增量映射方法将操作者手持控制器位姿映射到虚拟机器人 1。

(3) 采用人机解耦的信息物理映射模型, 同时采用基于虚拟现实空间的变比例位置增量映射方法将

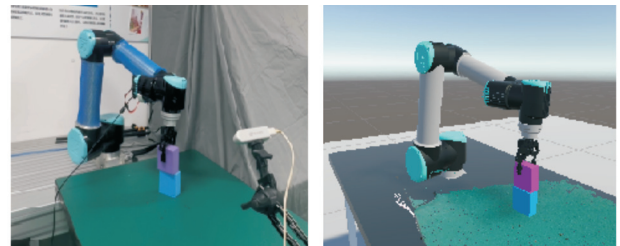


图 14 积木堆叠任务

Fig.14 Task of stacking building blocks

操作者手持控制器位姿映射到虚拟机器人 1.

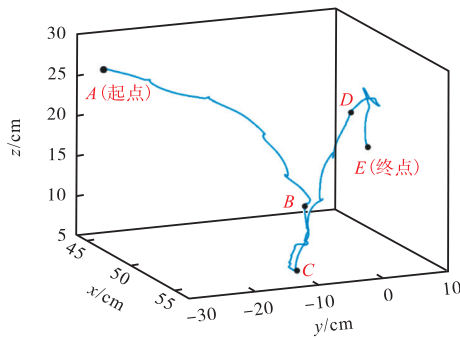
该实验招募了 15 名参与者,包括 14 男 1 女,年龄在 22~27 岁. 其中所有的参与者均没有使用虚拟现实设备的经验,并且均具备基础的机器人知识. 每位参与者将在 3 种条件下分别重复进行 5 次积木堆叠实验.

4.3 实验结果与分析

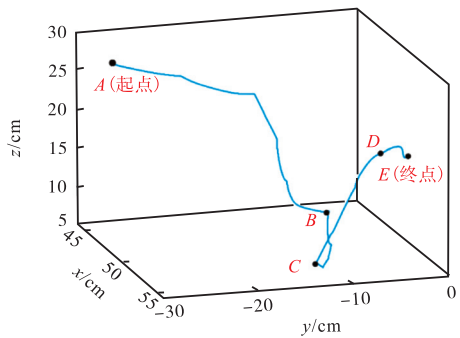
如图 15 所示,给出了其中一位参与者在 3 种不同条件下完成积木堆叠任务中虚拟机器人 1 的末端轨迹. 根据任务的不同阶段,可以将轨迹分为 4 个部分: A—B 段轨迹,操作者首先控制虚拟机器人 1 从 A 点开始向下运动,接近积木; B—C 段轨迹,操作者继续控制虚拟机器人 1 向下,并调整末端姿态对积木进行抓取; C—D 段轨迹,操作者控制虚拟机器人 1 向

上运动到另一积木的顶部附近; D—E 段轨迹,操作者控制虚拟机器人 1 调整积木姿态,将其准确地放置到另一积木上方,松开抓手实现积木堆叠. 由图 15(a)可以看出,操作者在条件 1 下完成任务的过程中因时延问题导致操作者手持控制器下达的指令与视觉反馈不同步,导致轨迹中出现多次方向的突变. 因为采用固定比例增量式映射法,在精确调整阶段需多次往返调整,影响实验效率. 图 15(b)、(c)中虚拟机器人 1 末端的轨迹更加平滑且不存在突变,说明人机解耦的信息物理映射模型消除了时延影响,提高了遥操作系统稳定性.

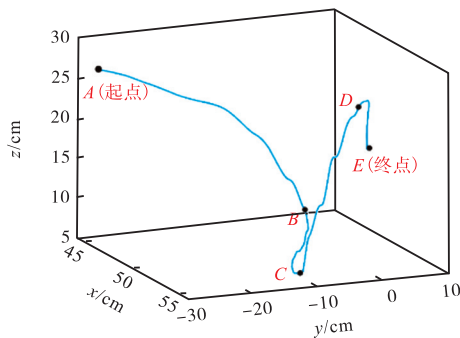
图 16 给出了 3 种条件下参与者完成积木堆叠任务的成功率和完成时间. 在 3 种条件下,操作者均表现出超过 60% 的成功率,这归功于基于点云与 Gen6D 位姿估计的场景重建方法为操作者提供了良好的临场感. 与条件 1 相比,操作者在条件 2 下没有大幅度提高实验的成功率,导致这一情况的原因是在操作的过程中出现过多次碰倒积木的情况. 但在条件 2 下操作者大幅度减少了操作时间,这说明人机解耦的信息物理映射模型消除了时延影响,无时延的遥操作提高了任务的完成效率. 此外,参与者在条件 3 下表现出更高的成功率,更短的完成时间以及最小的结果差异. 基于虚拟现实空间的变比例位置增量映射方法不仅大大提高了遥操作的灵活性,还避免了抓手与积木的碰撞.



(a) 条件 1



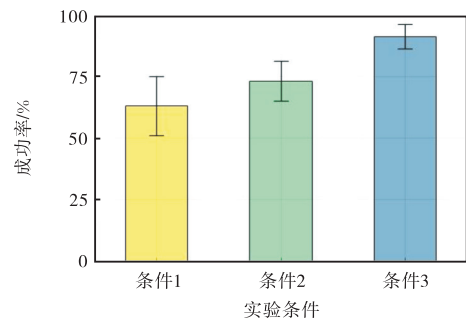
(b) 条件 2



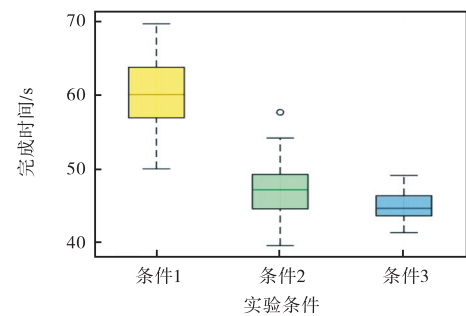
(c) 条件 3

图 15 虚拟机器人 1 末端在不同条件下的轨迹

Fig.15 Trajectories of the end of virtual robot 1 under different conditions



(a) 3 种条件下的实验成功率



(b) 3 种条件下的实验完成时间

图 16 参与者在实验中的表现

Fig.16 Performance of the participants in the experiment

表1为基于Gen6D位姿估计算法的场景重建策略和传统的基于点云实时传输的场景重建策略之间的实时性差异。实验以重建较低分辨率(640像素×480像素)的远程场景为例,来突出基于Gen6D位姿估计算法的场景重建策略对重建效率的提高。Gen6D位姿估计算法经过目标检测器、视点选择器、姿态精细器处理一帧RGB图像得到目标物体位姿的平均时间约92ms,且仅需向Unity传输18Byte的数据(目标物体的位姿 $[x, y, z, a, b, c]$),平均传输时间4ms,综合时延96ms。基于点云实时传输的场景重建策略合成一帧点云远程场景点云平均时间约62ms,但需要向Unity传输5893850Byte的数据,平均传输时间1155ms,综合时延1217ms。综上可以看出即使在低分辨率下,远程场景的点云数据量也是巨大的,这为点云的实时传输带来了巨大的负担。而基于Gen6D位姿估计的场景重建方法显著减少了通过网络传输的数据量,虽然平均处理时间相对于点云合成较长,但相比之下基于位姿估计结果用虚拟对象替换大量原始点云数据,大大减少了网络传输延迟,同时刷新率也能满足遥操作系统的实时性要求。

表1 Gen6D位姿估计算法与点云实时传输的实时性对比

Tab.1 Real-time performance comparison of Gen6D pose estimation algorithm with real-time transfer of point-cloud

| 场景重建策略 | 传输数据量/Byte | 处理时间/ms | 传输时间/ms | 综合时延/ms |
|-------------|------------|---------|---------|---------|
| 基于Gen6D位姿估计 | 18 | 92 | 4 | 96 |
| 基于点云实时传输 | 5 893 850 | 62 | 1 155 | 1 217 |

5 结 语

本文提出了一种基于人机解耦的信息物理映射模型的虚拟现实机器人遥操作系统,选用泛化性较强的Gen6D位姿估计算法以提高在非结构环境下的远程环境重建系统的适应能力。本文提出的人机解耦的信息物理映射模型为操作者提供了更直观的交互方式,消除了时延带来的影响,提高了操作的稳定性。同时本文提出的基于虚拟现实空间的变比例位置增量映射算法,根据虚拟空间中机器人末端与目标物体的距离改变位姿映射的比例,提高了操作效率与精度,并赋予虚拟现实遥操作系统避障的功能,提高了虚拟现实遥操作系统的安全性。

工作场景复杂和相机视线受到遮挡将会导致位姿估计的失败。在未来的研究中,笔者将侧重于提高本文所提出的虚拟现实遥操作系统遇到复杂工作环

境时的鲁棒性。同时,还考虑使用力反馈遥操作设备替代手持控制器,以便遥操作系统能完成更加精细的任务。

参考文献:

- [1] Hokayem P F, Spong M W. Bilateral teleoperation: An historical survey[J]. *Automatica*, 2006, 2(12): 2035-2057.
- [2] Lester D, Thronson H. Human space exploration and human spaceflight: Latency and the cognitive scale of the universe[J]. *Space Policy*, 2011, 27(2): 89-93.
- [3] Kot T, Novák P. Application of virtual reality in teleoperation of the military mobile robotic system TAROS[J]. *International Journal of Advanced Robotic Systems*, 2018, 15(1): 54819501.
- [4] 张莹. 人机交互遥操作机器人技术前沿和发展趋势——访东南大学宋爱国教授[J]. *微纳电子与智能制造*, 2022, 4(2): 1-3.
Zhang Ying. Frontier and development trend of human-computer interactive teleoperation robot technology—Interview with professor Song Aiguo of Southeast University[J]. *Micro-Nano Electronics and Intelligent Manufacturing*, 2022, 4(2): 1-3 (in Chinese).
- [5] Boessenkool H, Abbink D A, Heemskerk C J M, et al. Analysis of human-in-the-loop tele-operated maintenance inspection tasks using VR[J]. *Fusion Engineering and Design*, 2013, 88(9/10): 2164-2167.
- [6] Lee P S. The effects of camera perspective and field of view on performance in teleoperated navigation[C]// *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Los Angeles, USA, 2006, 50(16): 1528-1532.
- [7] Collett M, Chittka L, Collett T S. Spatial memory in insect navigation[J]. *Current Biology*, 2013, 23(17): 789-800.
- [8] DeJong B P, Colgate J E, Peshkin M A. Improving teleoperation: Reducing mental rotations and translations[C]// *IEEE International Conference on Robotics and Automation*. Los Angeles, USA, 2004, 4: 3708-3714.
- [9] Borges M, Symington A, Coltin B, et al. HTC Vive: Analysis and accuracy improvement[C]// *2018 IEEE RSJ International Conference on Intelligent Robots and Systems (IROS)*. Madrid, Spain, 2018: 2610-2615.
- [10] 王太勇, 于恩霖. 基于三维关键点投票的物体位姿估计方法[J]. *天津大学学报(自然科学与工程技术版)*,

- 2024, 57(3): 291-300.
- Wang Taiyong, Yu Enlin. Object pose estimation method based on 3D key points voting[J]. Journal of Tianjin University(Science and Technology), 2024, 57(3): 291-300(in Chinese).
- [11] Wonsick M, Padir T. Human-humanoid robot interaction through virtual reality interfaces[C]// 2021 IEEE Aerospace Conference (50100). Montana, USA, 2021: 1-7.
- [12] LeMasurier G, Allspaw J, Yanco H A. Semi-autonomous planning and visualization in virtual reality[EB/OL]. <https://arxiv.org/abs/2104.11827>, 2021-04-23.
- [13] Naceri A, Mazzanti D, Bimbo J, et al. The vicarios virtual reality interface for remote robotic teleoperation: Teleporting for intuitive tele-manipulation[J]. Journal of Intelligent & Robotic Systems, 2021, 101: 1-16.
- [14] Xu S Y, Moore S, Cosgun A. Shared-control robotic manipulation in virtual reality[C]//2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA). Ankara, Turkey, 2022: 1-6.
- [15] Wang C, Cho Y K. Smart scanning and near real-time 3D surface modeling of dynamic construction equipment from a point cloud[J]. Automation in Construction, 2015, 49: 239-249.
- [16] Rosen E, Whitney D, Phillips E, et al. Testing robot teleoperation using a virtual reality interface with ROS reality[C]//Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI(VAM-HRI). Chicago, USA, 2018: 1-4.
- [17] Lipton J I, Fay A J, Rus D. Baxter's homunculus: Virtual reality spaces for teleoperation in manufacturing[J]. IEEE Robotics and Automation Letters, 2017, 3(1): 179-186.
- [18] Zhou T, Zhu Q, Du J. Intuitive robot teleoperation for civil engineering operations with virtual reality and deep learning scene reconstruction[J]. Advanced Engineering Informatics, 2020, 46: 101170.
- [19] Wonsick M, Padir T. A systematic review of virtual reality interfaces for controlling and interacting with robots[J]. Applied Sciences, 2020, 10(24): 9051.
- [20] Liu Y, Wen Y, Peng S, et al. Gen6D: Generalizable model-free 6-DoF object pose estimation from RGB images[C]// European Conference on Computer Vision. Tel Aviv, Israel, 2022: 298-315.

(责任编辑: 王晓燕)