doi:10.6041/j.issn.1000-1298.2025.10.070

面向无模型物体的机器人视觉-力觉协同交互方法

黄玲涛　王琳轩　孔紫静　张红彦
（吉林大学机械与航空航天工程学院，长春 130022）

摘要：视觉-力觉多模态融合是提升机器人操作性能的核心技术，针对当前视觉定位方法对高质量三维模型的依赖性及机器人与环境之间交互能力不足等问题，本文提出了一种面向无模型物体的机器人视觉-力觉协同交互方法。采用 Colmap 三维重建算法获取物体点云数据，通过位姿估计技术实现无模型物体定位。针对视觉硬件精度不足导致的定位偏差问题，采用基于力觉的物体交互策略，利用决策树算法与柔顺控制策略的动态调整机器人运动。最后，进行了物体抓取堆叠和约束放置实验。实验结果表明，所提出的视觉定位方法能够实现无模型物体定位，在未经建模物体定位方面展现出良好的泛化性能；基于力觉的物体交互策略能够进行实时碰撞预测和调整，调整后约束放置成功率从 53% 提升至 87.5%，显著提升了多目标受限空间装配作业任务可靠性。

关键词：机器人；无模型物体；视觉和力觉；交互技术；机器学习
中图分类号：TP242　　**文献标识码**：A　　**文章编号**：1000-1298(2025)10-0770-10　　**OSID**：

Vision – Force Collaborative Interaction Method for Model-free Objects in Robotic Systems

HUANG Lingtao　WANG Linxuan　KONG Zijing　ZHANG Hongyan
(*School of Mechanical and Aerospace Engineering*, *Jilin University*, *Changchun* 130022, *China*)

Abstract: Aiming to achieve visual localization of model-free objects and enhance the interaction capabilities between robots and their environment, a collaborative interaction method that integrated robot vision and haptics specifically tailored for model-free objects was proposed. The approach utilized the Colmap 3D reconstruction algorithm to obtain point cloud data of objects and employs pose estimation techniques to localize these model-free objects. To address localization biases arising from the insufficient accuracy of visual hardware, a force-based interaction strategy was adopted, which dynamically adjusted the robot's movements by using decision tree algorithms and compliance control strategies. Furthermore, experiments involving object grasping, stacking, and constrained placement were conducted to evaluate the method's effectiveness. The experimental results indicated that the proposed visual localization technique successfully localized model-free objects, demonstrating strong generalization capabilities in the localization of unmodeled objects. Additionally, integrating real-time collision prediction and adjustment based on decision tree algorithms and compliance control strategies significantly enhanced the success rate of multi-target assembly tasks within constrained environments. This research contributed to advancing the field of human-robot interaction, particularly in scenarios that involved unmodeled objects, thereby facilitating more intuitive and effective robotic manipulation in real-world applications. Overall, the outlined vision – force collaborative interaction method showed promise for practical implementation in dynamic environments, paving the way for future developments in robotic systems that can effectively engage with their surroundings and adapt to various tasks involving model-free objects.

Key words: robot; model-free objects; vision and haptics; interaction technology; machine learning

收稿日期：2025-04-11　修回日期：2025-05-27
基金项目：吉林省重点研发计划项目(20200401130GX)
作者简介：黄玲涛(1979—)，男，副教授，主要从事机器人控制和主从控制及力反馈技术研究，E-mail：hlt@jlu.edu.cn
通信作者：张红彦(1978—)，女，副教授，主要从事机器人技术和机器视觉处理研究，E-mail：Zhanghy@jlu.edu.cn

0　引言

工业 5.0 推动智能制造向人机协作方向转型，要求机器人具备非结构化环境下操作未知无模型物体能力[1]。传统依赖预定义三维模型的交互方法，如模板匹配或预训练深度学习模型，难以适应动态交互场景，对物体几何先验的依赖性较强，难以泛化到现实中的其他物体，限制了其在真实工业环境中的应用。多模态融合与智能交互成为提升机器人操作可靠性的核心技术。传统单模态感知方法在实际应用中存在显著局限性，例如视觉定位易受环境干扰，而力觉控制在初始定位阶段效率不足。因此，视觉-力觉多模态融合技术的发展为其提供了解决方案。

位姿估计是计算机视觉中的关键技术，在视觉定位方面具有重要应用价值。基于模板匹配的方法，通过预建姿态模板库实现位姿估计，适用于弱纹理目标，具有较好的适应性[2]。Linemod 算法通过 Sobel 算子提取 RGB 图像的梯度幅值特征，通过线性存储结构优化了在线匹配效率[3]。改进型 CT - Linemod 算法进一步提升了匹配性能[4]。基于投票的方法分为直接投票和间接投票。文献[5]提出了 PPF 算法利用点云中点对的距离和法线特征生成 PPF 描述子并构建哈希表，通过检索匹配点对并直接投票确定最优位姿。间接投票通过局部特征匹配和投票机制计算位姿，使用特征点对齐提升位姿估计精度[6]。深度学习技术为位姿估计提供了新的解决方案，PointNet 通过二维卷积提取特征，显著减少了计算量[7]。SSD - 6D 通过分解姿态空间降低训练复杂度，从而有效处理对称物体[8]。然而，上述位姿估计方法仅适用于经过特定训练的物体，难以泛化到实际无模型物体，因此，需要对无模型物体的视觉定位方法开展研究。此外，视觉硬件精度不足会导致物体定位存在误差，需要使用力觉感知技术提升机器人与环境的交互能力。

力觉感知技术通过实时感知接触力为控制系统提供反馈，在工业机器人柔顺控制领域发挥重要作用。力觉柔顺控制分为主动柔顺[9-10]和被动柔顺[11-12]2 类。被动柔顺控制通过弹簧等机构实现机器人与环境的适应性接触，但难以满足高精度需求。主动柔顺控制领域，文献[13]提出外部接触力位姿随动算法，提升了柔顺姿态调整和可控装配接触应力。文献[14]设计了一种利用 Maxwell 模型的导纳控制策略，为机器人末端执行器提供先进的塑性柔顺性。基于学习的力控技术也取得显著进展[15-16]。

近年来，视觉-力觉多模态融合技术已取得了显著进展[17-19]。现有方法仍存在显著局限，包括依赖大量仿真/真实数据训练模型，调参复杂度高，实时性差。针对现有视觉力觉融合方法依赖精确建模的问题，同时为了提升机器人与无模型物体和环境之间的交互能力，本文提出一种面向无模型物体的机器人视觉-力觉协同交互方法，包括无模型物体视觉定位方法和基于力觉的物体交互策略 2 部分。视觉定位方法依托于 Colmap 三维重建算法获取目标物体的点云数据并进行预处理，通过 Gen6D 网络处理实时采集的图像，完成无模型物体定位任务。针对视觉硬件精度不足导致的定位偏差的问题，引入基于力觉的物体交互策略，通过决策树算法与柔顺控制策略动态调整机器人运动，以提升机器人与环境间的交互能力。

1　系统整体组成

机器人视觉-力觉协同交互系统框图如图 1 所示。系统由视觉感知模块、力觉感知模块和机器人控制模块组成。其中，视觉感知模块通过 Kinect 相机采集图像并进行 Colmap 三维重建，实现对无模型物体的视觉定位。力觉感知模块通过 Leprino 力传感器感知机器人与环境之间的接触力，结合决策树和导纳控制进行力觉预测调整。机器人控制模块控制 Epson 机器人完成物体抓取及约束放置等操作任务。系统有 2 种工作模式：①视觉模式，通过三维重建及位姿估计实现对无模型物体视觉定位，同时采用柔顺控制方法对物体位置进行辅助调整。②视觉-力觉模式，在视觉模式基础上增加决策树模型预测方法。

2　无模型物体视觉定位

位姿估计是实现物体定位的关键技术，现有方法多局限于特定对象或类别，少数泛化方法依赖高质量 3D 模型。为此，本文提出一种无模型物体的视觉定位方法，为后续机器人操作任务提供支持。该方法由三维重建、位姿估计和基于深度信息的定位 3 部分组成。

2.1　Colmap 三维点云重建

在计算机视觉中，Colmap 算法常被用作提取关键点、匹配特征和构建稀疏三维点云。Colmap 三维重建主要基于 2 个原理，即运动恢复结构（SfM[20]）和多视图立体匹配（MVS[21]）。SfM 通过重叠图像的投影恢复场景中的三维结构和相机位姿。输入是 1 组从不同视角拍摄的同一物体图像，输出是场景三维重建和所有图像的内外参数。MVS 利用 SfM

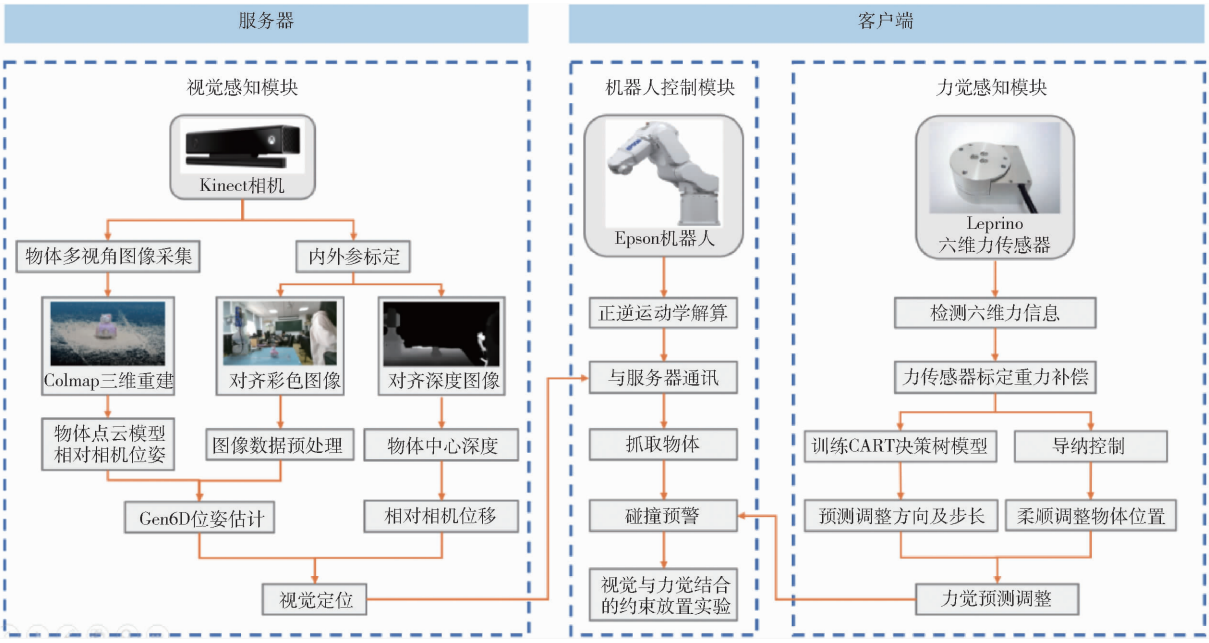
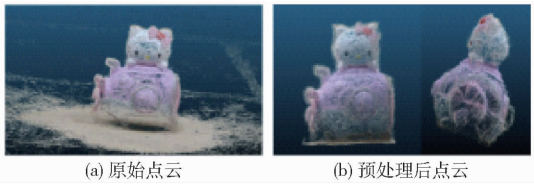


图 1 系统整体框图

Fig. 1 Block diagram of overall system

的输出,为每个像素计算深度和法线信息并生成场景稠密点云。选择 Colmap 三维重建算法进行数据准备工作。

首先图像数据采集。使用手机或相机围绕目标物体拍摄一圈,每幅图像拍摄角度不超过 30°,以确保相邻图像之间有足够的重叠区域,避免因角度过大而导致图像信息丢失或重建错误,保证三维重建结果准确性。共采集 371 幅图像。对采集到的图像进行三维重建。在服务器端部署 Colmap 算法,重建的信息包括场景点云数据以及每幅图像相机内外参数,这些参数包含相机位置、姿态、焦距等信息。随后,点云数据预处理。由于三维重建出的原始点云包含周围环境点云信息,这些环境点云可能会干扰后续位姿估计,因此需要对原始点云进行裁剪和滤波等处理,处理后点云结果如图 2 所示。最后,确定点云坐标系。选择物体朝向作为 X 轴正方向,将物体底面拟合为平面,该平面法向量定义为 Z 轴正方向,Y 轴正方向通过右手定则确定。



(a) 原始点云 (b) 预处理后点云

图 2 目标物体点云

Fig. 2 Target object point clouds

2.2 Gen6D 物体位姿估计

Gen6D 网络整体结构主要由对象检测器、视点选择器和姿态细化器组成[22]。检测物体中心二维投影 q,并估计包围单位球体紧凑方形边界框 S_q。物体中心深度 d 计算式为

$$d = 2f/S_q \tag{1}$$

式中 f——虚拟焦距

通过调整主点投影 q,可以确定物体中心位置,从而为物体位姿提供平移信息。该方法无需高质量三维模型,不依赖预训练参数与深度图像数据。

在服务器端部署 Gen6D 网络并配置物体点云模型,通过相机获取场景图像并输入网络处理,即可实现目标物体位姿估计。网络所定位的目标物体被三维立体框包围,但仅依靠立体框难以分辨物体旋转角情况。为实现物体坐标轴可视化,首先,选定物体三维立体框顶点,计算出该立体框长宽高(l,w,h),由上顶点向中心偏移 $l/2$、$w/2$、$h/2$,得到三维立体框中心点;最后,使用黄色、绿色和红色粗实线表示 X 轴、Y 轴和 Z 轴正方向。在保持物体位置不变的前提下,将物体分别绕 Z 轴旋转 90°,物体位姿估计效果如图 3 所示,可以观察到物体朝向变化。

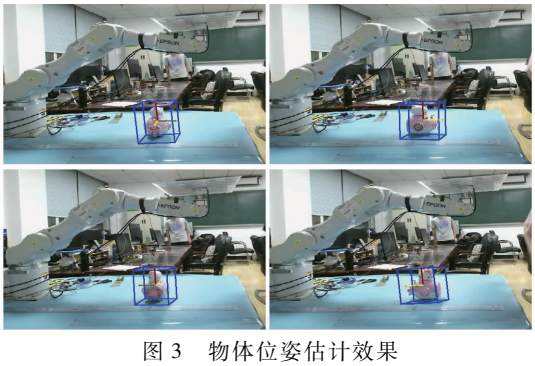


图 3 物体位姿估计效果

Fig. 3 Object pose estimation performance

Colmap 生成的三维模型可通过相机视角与特

征点之间的几何关系进行重建。由于其点云模型缺乏物理单位，无法直接对应实际物体尺寸。为获得公制尺度（如 m、cm 等），需将重建结果转换至标准尺度，步骤为：

（1）确定尺度校正比例。选取物体上 2 个特征点，并记录其实际距离。利用点云可视化软件（CloudCompare）载入点云数据，定位所选特征点。运用欧氏距离公式计算模型两点间距。通过计算结果与实际尺寸比值，确定尺度转换系数。

（2）转换相机位姿信息。读取 Colmap 输出的 images. txt 文件，提取每幅图像平移向量。将每幅图像平移向量乘以尺度校正比例，从而得到以米为单位的平移向量。

（3）对点云进行缩放对齐。model_aligner 是 Colmap 中用于地理配准的工具。使用 model_aligner 对重建的点云模型进行缩放操作，并将其与目标坐标系进行对齐。

同时，由于原 Gen6D 算法将物体点云模型归一化至单位球内，并将物体从 Colmap 坐标系旋转到指定坐标系。因此，在求解真实位姿时，需对物体进行反旋转和反缩放操作。

采用 $\boldsymbol{X}_{_col}$、$\boldsymbol{X}_{_cam}$ 和 $\boldsymbol{X}_{_norm}$ 表示物体点云坐标系、相机坐标系和归一化后的点云坐标系。（$\boldsymbol{R}_{_col}$，$\boldsymbol{T}_{_col}$）为网络输出的物体位姿，即 $\boldsymbol{X}_{_col}$ 相对于 $\boldsymbol{X}_{_cam}$ 的变换矩阵；（$\boldsymbol{R}_{_norm}$，$\boldsymbol{T}_{_norm}$）为 $\boldsymbol{X}_{_norm}$ 相对于 $\boldsymbol{X}_{_cam}$ 的变换矩阵，其中

$$\boldsymbol{X}_{_cam} = \boldsymbol{R}_{_col}\boldsymbol{X}_{_col} + \boldsymbol{T}_{_col} \tag{2}$$

其中

$$\boldsymbol{X}_{_col} = (\boldsymbol{R}_{_col}, \boldsymbol{T}_{_col}) \tag{3}$$

$\boldsymbol{X}_{_col}$ 通过旋转矩阵 $\boldsymbol{R}_{_norm}$ 和缩放因子 $S_{_norm}$ 转换到 $\boldsymbol{X}_{_norm}$，表达式为

$$\boldsymbol{X}_{_norm} = S_{_norm}(\boldsymbol{R}_{_norm}\boldsymbol{X}_{_col}) \tag{4}$$

（$\boldsymbol{R}_{_pre}$，$\boldsymbol{T}_{_pre}$）为 Gen6D 网络输出物体位姿，$\boldsymbol{X}_{_norm}$ 相对于 $\boldsymbol{X}_{_cam}$ 的变换矩阵，即

$$\boldsymbol{X}_{_cam} = \boldsymbol{R}_{_pre}\boldsymbol{X}_{_norm} + \boldsymbol{T}_{_pre} \tag{5}$$

将 $\boldsymbol{X}_{_norm}$ 代入式（4），得出

$$\boldsymbol{X}_{_cam} = \boldsymbol{R}_{_pre}\boldsymbol{R}_{_norm}\boldsymbol{X}_{_col} + \boldsymbol{T}_{_pre}/S_{_norm} \tag{6}$$

所求的（$\boldsymbol{R}_{_col}$，$\boldsymbol{T}_{_col}$）表达式为

$$\boldsymbol{R}_{_col} = (\boldsymbol{R}_{_pre}\boldsymbol{R}_{_norm}) \tag{7}$$

$$\boldsymbol{T}_{_col} = \boldsymbol{T}_{_pre}/S_{_norm} \tag{8}$$

按照式（7）、（8）对（$\boldsymbol{R}_{_pre}$，$\boldsymbol{T}_{_pre}$）进行处理，获得世界坐标系下物体实际的（$\boldsymbol{R}_{_col}$，$\boldsymbol{T}_{_col}$）。

用 $\{O\}$、$\{C\}$ 和 $\{B\}$ 表示物体坐标系、相机坐标系和机器人坐标系，$\{O\}$ 在坐标系 $\{B\}$ 下变换矩阵为 ${}^{B}_{O}\boldsymbol{T}$ 为

$${}^{B}_{O}\boldsymbol{T} = {}^{B}_{C}\boldsymbol{T}\,{}^{C}_{O}\boldsymbol{T} \tag{9}$$

${}^{C}_{O}\boldsymbol{T}$ 为物体位姿矩阵，由 Gen6D 网络求出；${}^{B}_{C}\boldsymbol{T}$ 为 $\{C\}$ 相对于 $\{B\}$ 的外参矩阵，由相机外参标定获得。由式（9）即可求出物体相对于机器人基坐标系下位置。

为验证 Gen6D 网络定位精度，将物体随机放置在实验台 8 个不同位置，并采集图像送入网络进行处理，提取输出旋转矩阵中第 4 列平移向量作为测量值。同时人工标注出真实值，使用激光测距仪分别测量出物体距机器人基座坐标系 X 和 Y 轴方向的真实距离，由于物体摆放在实验台上，其高度一半即为 Z 轴方向真实值。经计算，物体平均位移误差为 3. 697 mm，误差主要来源于 RGB 图像分辨率及算法精度。

2.3　基于深度信息的物体定位方法

在机器人操作过程中，物体结构特征可能会发生变化。以序列化装箱作业为例，随着装载物体数量增加，箱体几何结构特征会发生变化。此时，2. 2 节中所述的定位方法可能会产生较大的误差。尽管物体中心投影点保持不变，但由于边框发生变化，导致式（1）中 d 发生变化，从而影响定位精度。为解决这一问题，本文提出了一种基于深度信息的物体定位方法。该方法首先利用 2. 2 节中的 Gen6D 物体位姿估计获取目标物体中心点在图像平面的投影像素坐标，通过相机内参标定矩阵建立 RGB－D 像素映射关系，从而获取物体中心点在相机坐标系下深度。采用逆投影变换将图像坐标转换为相机坐标系下的三维坐标（x，y，z）。最后，通过相机外参矩阵，将物体的三维坐标转换至机器人基坐标系下。

为了验证基于深度信息的物体定位方法精度，对在机器人坐标系中的物体位置误差进行了检测。使用激光测距仪分别测量物体在 X、Y 轴方向位移，并将这些测量结果与基于深度信息的物体定位方法得到的物体位移进行对比。在实验中，物体 Z 轴高度被固定为 50 mm。对物体在不同位置下进行了 12 次独立测量，真实值与测量值对比结果如图 4 所示。通过计算，物体平均位移误差为 4. 121 mm，X、Y、Z 轴方向均无异常值（1. 5IQR 内），说明测量数据与真实值基本一致。与 2. 2 节中提出的定位方法误差相比，基于深度信息的定位方法误差较大，主要原因在于该方法依赖于深度图像，而 Kinect 相机深度图像分辨率较低（512 像素 × 424 像素），导致在测量物体边缘或细节时的精度较低。此外，深度图像中的噪声也会增加定位误差。

为了验证基于深度信息的物体定位方法在人机交互中的应用效果，需引入具有结构特征变化的物体装配作业任务，以评估其有效性。另外，由于视觉

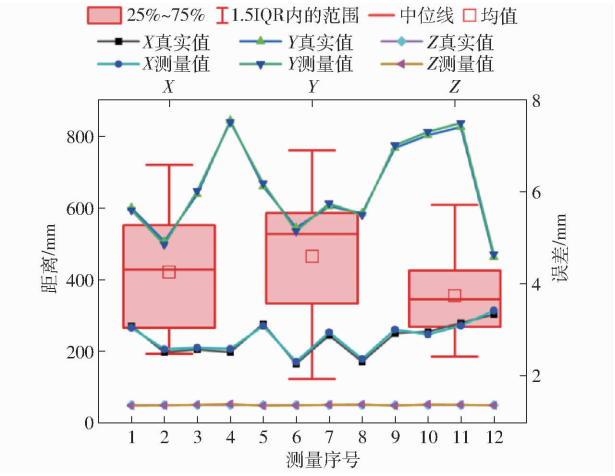


图 4 真实值与测量值对比

Fig. 4 Comparison of true values and measured values

定位过程中可能存在一定误差,因此,需通过受限空间中的多目标装配作业任务验证所提出的无模型物体视觉定位方法的可靠性。

3 基于力觉的物体交互策略

由于硬件设备性能及算法局限性,视觉定位结果仍存在 3 ~4 mm 误差,这可能导致机器人执行任务时发生失败。为提高机器人与环境之间的交互能力,本文提出了一种基于力觉的物体交互策略。该策略包括 3 个关键环节:力觉感知、决策与调整以及柔顺控制。力觉感知通过安装在机器人末端执行器上的力觉传感器,实时获取接触力信息[23],决策与调整基于力觉信息判断是否发生碰撞,若碰撞发生,使用机器学习决策树模型预测交互信息,并调整运动方向和步长。柔顺控制通过柔顺控制策略,使机器人在与物体接触时表现出适当的顺应性,避免刚性碰撞引起损坏或误差。

3.1 CART 决策树交互信息预测

机器学习通过对大量样本数据的深度挖掘,能够有效提取潜在规律并实现精准预测。在装配作业任务中,被装配物体与纸箱边缘不同位置接触时,六维力信息大小和方向存在显著差异。因此,本文以六维力数据作为特征量,接触点偏移方向和距离作为目标变量,训练机器学习模型。

S_{x+} 表示 X 正方向偏移距离, S_{y-} 表示 Y 负方向偏移距离。将纸箱内部划分为 6 个等面积区域,记录左侧中心点位置,其他区域中心通过偏移计算得到。数据采集规则包括:①确定采集起点:将物体移动至目标区域正上方,确保物体与区域对齐,记录此位置为起始点,将(S_x,S_y)记为(0,0)。②基于机器人坐标系,依次向 $X+$、$X-$、$Y+$、$Y-$ 4 个方向移动并采集数据。③偏移距离:控制机器人每次向正方向偏移 2 mm,以确保力与力矩变化完整记录。

在实际力数据采集场景中,机械手爪根据预设程序依次执行抓取动作。设定的机器人末端运动轨迹为:机器人移动至区域中心点,偏移至某一方向,向下移动至接触物体,采集接触时的力数据。机器人上升并偏移至下一方向,重复上述过程。该过程循环多次,以确保在不同方向和位置下均能有效采集力数据。为了获取足够的样本数据,机器人在 1 次采集中沿正方向偏移 16 次,每次下降后采集 15 组数据,共计采集 4 个方向的力数据,最终获得 7 680 组训练数据。由于机器人末端运动较为缓慢,在力采集过程中,惯性力可忽略不计。

图 5 为 $Y-$方向下采集的部分六维力数据。由图 5 可知,$Y-$方向移动时,随着偏移距离 S_y 逐渐增加,F_x、F_y 和 F_z 逐渐增大,表明可使用机器学习模型预测步长。同时,不同偏移方向对应的力具有正负差异。在图 5 中,物体向 $Y-$方向移动时,F_x 和 F_y 为正值,F_z 为负值,且随着 S_y 增加,F_x 和 F_y 呈上升趋势,表明可以使用机器学习模型预测调整方向。

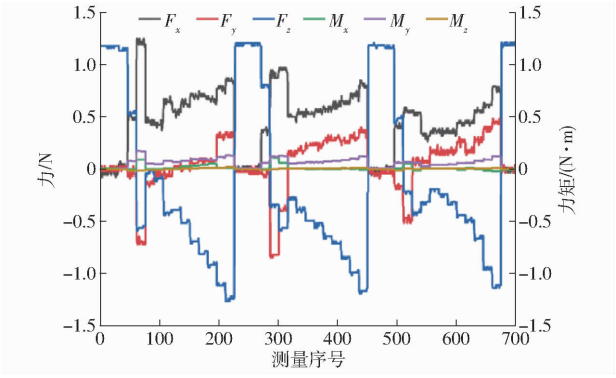


图 5 Y-方向采集力数据

Fig. 5 Y-axis force data collection

机器学习算法根据学习目标分为分类和回归 2 类,其中分类用于预测类别标签,回归用于预测数值变量。采用回归模型预测调整方向和步长,目标特征为连续变量。定义平均误差 S_{mse} 为

$$S_{mse}=\frac{1}{n}\sum_{i=1}^{n}(\hat{S}_i-S_i)^2 \tag{10}$$

式中 $\hat{S}_i$——预测值,mm

S_i——真实值,mm

S_{mse} 通过平方误差实现对较大误差的加权惩罚,且对异常值敏感。对比 KNN、支持向量机、线性回归、XGBoost、Catboost、GBDT、梯度提升、Ridge、Lasso 及决策树这 10 种机器学习模型,并在未调参情况下进行预测实验,实验结果如图 6 所示。由图 6 可知,决策树表现最优误差最小,因此本文选择决策树模型进行预测。

将所采集六维力数据集按照比例 8:2 分为训练

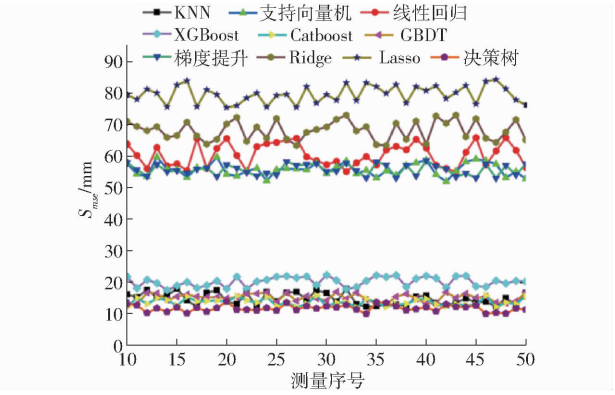


图 6　不同模型性能表现

Fig. 6　Performance of different models

集和测试集，并依次进行特征选择、决策树生长以及剪枝优化，最终确定决策树 3 个核心超参数：树最大深度、最小样本数及叶节点最小样本数。为寻找参数空间中最佳参数组合，后续参数在前一参数最优值的基础上搜索。参数优化结果如图 7 所示，当树最大深度取 25 时，S_{mse} 最小，且随着树深度增加，S_{mse} 逐渐稳定，因此选择最大深度为 25；最小样本数取 2 时，误差最小，进一步结合叶节点的最小样本数为 1 时，误差最小。最终，CART 决策树模型最优参数组合为最大深度 25、最小样本数 2 和叶节点最小样本数 1，模型均方误差为 3.842 1 mm。采用决定系数 R^2 作为模型性能评价指标，R^2 越接近 1，表示模型解释能力越强。通过对决策树模型进行测试评估，结果显示，本文模型在训练集和测试集上 R^2 分别为 0.954 和 0.928，表明模型具有良好的稳健性和泛化能力。

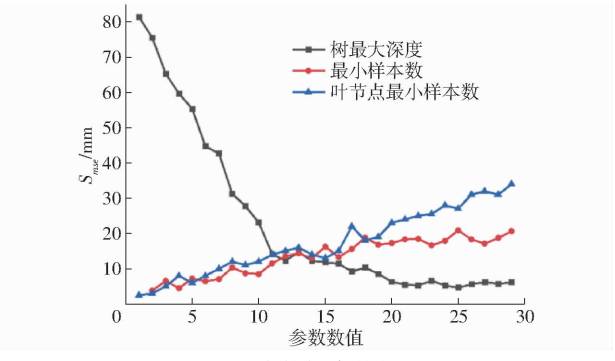


图 7　决策树参数优化

Fig. 7　Decision tree parameter optimization

3.2　基于力觉的物体交互策略

机器人柔顺控制通过实时调整机器人运动参数，确保其与环境交互过程中的柔性适应能力。在传统位置控制模式下，机器人接触物体时会产生较大接触力，导致交互任务失败。为此，采用基于导纳原理的柔顺控制策略，使机器人在执行任务时具备顺从性，其导纳控制结构如图 8 所示。

机器人末端位姿记为 $\boldsymbol{X}$，其位姿变化可表示为

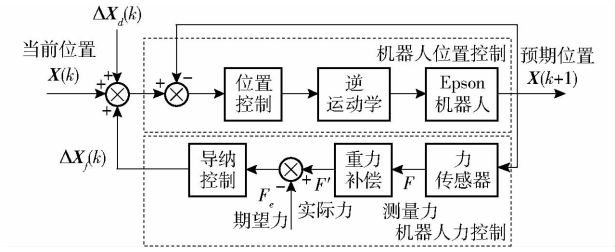


图 8　导纳控制结构图

Fig. 8　Admittance control diagram

$$\boldsymbol{X}(k+1)=\boldsymbol{X}(k)+\Delta\boldsymbol{X}_d(k)+\Delta\boldsymbol{X}_f(k) \tag{11}$$

式中　$\boldsymbol{X}(k+1)$——预期位置向量

$\boldsymbol{X}(k)$——当前位置向量

$\Delta\boldsymbol{X}_d(k)$——位控方向位置增量

$\Delta\boldsymbol{X}_f(k)$——力控方向位置增量

在执行操作任务时，机器人依据环境几何特性设定力控与位置控制方向。通过引入对角矩阵 $\boldsymbol{P}$ 控制机器人力控和位控方向，此时式(11)可整理为

$$\boldsymbol{X}(k+1)=\boldsymbol{X}(k)+(\boldsymbol{I}-\boldsymbol{P})\Delta\boldsymbol{X}_d(k)+\boldsymbol{P}\Delta\boldsymbol{X}_f(k) \tag{12}$$

式中　$\boldsymbol{I}$——6×6 单位矩阵

$\boldsymbol{P}$——6×6 对角矩阵

当机器人末端不受力与力矩的控制时，$\boldsymbol{P}$ 中元素为 0；受力与力矩的控制时，$\boldsymbol{P}$ 中对应对角元素为 1。机器人位姿调整量由环境接触力决定，系统通过设定位移反馈系数将接触力偏差映射为位置补偿量。力控方向位置增量为

$$\Delta\boldsymbol{X}_f(k)=K_1(\boldsymbol{F}'-\boldsymbol{F}_e) \tag{13}$$

式中　K_1——反馈系数，取 0.1 mm/N[23]

$\boldsymbol{F}'$——实际力与力矩

$\boldsymbol{F}_e$——期望力与力矩

在无力控与有力控 2 种模式下进行物体推动实验。在无力控模式下，采用位置控制方法；在有力控制模式下，应用力与力矩控制，即柔顺控制策略。实验以饮料盒为对象，机械手沿 X 轴推动至目标位置(220 mm)。通过对比 2 种控制模式下推动效果，评估基于导纳的柔顺控制方法在实际应用中的性能。

有力控模式：机器人通过力传感器实时检测 X 轴方向的实际接触力 $\boldsymbol{F}'$、$\boldsymbol{F}_e$ 则通过预先实验进行确定。具体而言，在有力控模式下进行 X 轴方向柔顺推动实验时，经重复测试确定物体到达目标位置的标志是 X 轴方向阈值为 0.5 N。因此，力控实验中设定 X 轴方向期望力 $\boldsymbol{F}_e$ 为 0.5 N，但考虑到力传感器测量有误差，选择将期望力变化区间设为 ±0.05 N。接触力与期望力偏差映射为位置补偿量，控制机器人柔顺地推动物体至目标位置。力和位移变化曲线如图 9a 所示。由图 9 可知，0.5 s 时，机械手爪接触饮料盒，接触力突变，推动过程中力在 0 N 附近波

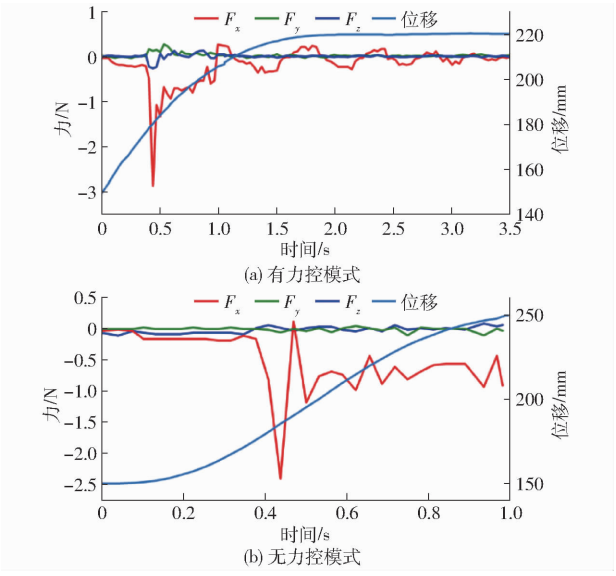


图 9 力和位移变化曲线

Fig. 9 Force-displacement curve

动,最终目标位置 220 mm 处停止,并且力为 0。

无力控模式:力和位移变化曲线如图 9b 所示。0.4 s 处,机械手爪接触饮料盒,接触力发生突变,F_x 在 -0.75 N 左右波动,推动过程不平稳且冲击力较大,并且饮料盒未能在期望位置停止。

实验结果表明,本文提出的柔顺控制策略在物体推动任务中表现出良好的性能,系统能够平稳地将目标物体推至预定位置,与传统的位置控制相比,柔顺控制有效降低了接触冲击。因此,在后续实验中使用柔顺控制策略校正物体位置。

4 实验

4.1 机器人实验平台

实验平台如图 10 所示,主要包括 Epson 六自由度机器人、Kinect 型相机、Leptrino 六维力传感器及服务器端和客户端主机。服务器端主机负责处理 Kinect 型相机采集的图像信息,实现无模型目标物体视觉定位,将物体定位信息发送至客户端主机。客户端主机负责接收物体坐标信息,并实时获取 Leptrino 六维力传感器的力信息,控制机器人的运动,完成作业任务。Kinect 型相机和六维力传感器通过 USB 3.0 接口与服务器主机相连,客户端和服务器端主机通过 TCP 进行通讯传输物体位置信息。

4.2 相机标定

采用“眼在手外”的相机固定模式建立标定系统。在该模式下,标定板被固定在机器人第六轴末端的法兰盘上,并通过调整机器人姿态,采集 20 幅标定图像。利用这些标定图像以及对应的机器人关节角度信息,进行外参计算[24],最终得到相机外参

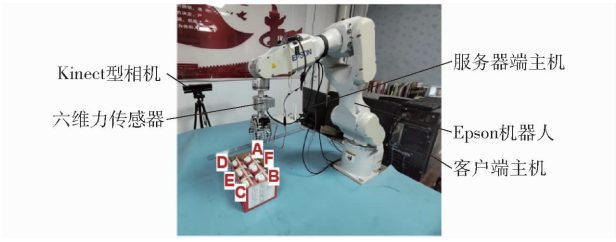


图 10 实验平台

Fig. 10 Experimental platform

矩阵为

$$ {}^{B}_{C}\boldsymbol{T} = \begin{bmatrix} -0.069\,7 & -0.006\,4 & -0.997\,5 & 1.127\,4 \\ 0.955\,1 & -0.070\,8 & -0.069\,1 & 0.804\,8 \\ -0.070\,2 & -0.997\,5 & 0.011\,3 & 0.360\,4 \\ 0 & 0 & 0 & 1 \end{bmatrix} $$

为了实现彩色图像与深度图像对齐,需要进行相机内参标定[25]。采集同一时刻和同一场景下彩色图像和红外图像,分别求得彩色图像和深度图像内参,结果为

$$ \boldsymbol{K}_c = \begin{bmatrix} 1\,059.4 & 0 & 964.6 \\ 0 & 1\,059.9 & 526.7 \\ 0 & 0 & 1 \end{bmatrix} $$

$$ \boldsymbol{K}_d = \begin{bmatrix} 517.79 & 0 & 265.10 \\ 0 & 495.06 & 207.14 \\ 0 & 0 & 1 \end{bmatrix} $$

利用 20 幅标定图像计算重投影误差,其中彩色相机平均重投影误差为 0.016 像素,深度相机平均重投影误差为 0.69 像素。误差只小于 1 像素,表示相机内参标定精度较高。

4.3 抓取堆叠实验

物体视觉定位是实现人机交互的关键技术,其精度直接影响物体检测与定位性能。为验证本文提出的无模型物体视觉定位方法在动态环境中目标识别与定位能力,设计了机器人抓取堆叠实验。实验使用 2 个标准杯子,通过 Gen6D 网络实时检测其空间位置,驱动机器人完成抓取-堆叠任务。任务完成后,操作人员随机调整杯子位置,系统需自主识别并更新目标位置,判断是否继续堆叠。

抓取堆叠实验流程为:系统通过视觉算法定位物体 1 和物体 2 空间位置,控制机器人执行抓取-移动-堆叠操作。为保证安全性,建立了基于力传感器的保护机制。当 Z 轴力发生突变时,系统立即停止下行动作,防止碰撞风险。实际场景中的物体抓取堆叠实验如图 11 所示。

为减少人为因素对实验随机性的影响,邀请 4 名操作人员独立完成 20 组实验,每组包括 3 次堆叠。在每组实验中,操作人员负责随机调整物体 1 或物体 2 的位置 3 次,系统通过视觉算法定位物体

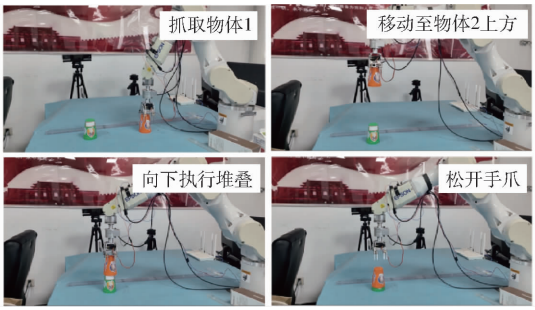


图 11　物体抓取堆叠实验

Fig. 11　Object grasping and stacking experiment

1 和物体 2 的空间位置，控制机器人自主完成抓取-移动-堆叠操作。若 3 次堆叠均成功，则该组实验成功；若发生碰撞或系统未能及时追踪新位置，则为失败。实验结果见表 1，平均成功率为 92.5%，堆叠成功率为 96.2%，少量失败由机械手爪夹持力度不足引起。

表 1　物体抓取堆叠实验结果

Tab. 1　Experimental results of object grasping and stacking

编号	成功次数/总次数	成功率/%	堆叠成功次数/堆叠次数	堆叠成功率/%
1	19/20	95	59/60	98.3
2	18/20	90	57/60	95.0
3	18/20	90	57/60	95.0
4	19/20	95	58/60	96.6

由表 1 可知，基于 Gen6D 的定位方法能够对无模型物体进行定位，并可靠地完成抓取及堆叠任务，具有较高的成功率。因为实验中的大口径杯子具有较高容错性，该方法仍需在高精度任务中进一步验证有效性。

4.4　约束放置实验

设计了具有严格放置容差的物体约束放置实验，通过对比视觉模式和视觉-力觉模式的操作成功率，探究 2 种模式在受限空间交互中的可靠性。实验中使用 6 个相同规格的饮料盒，尺寸为 70 mm × 70 mm × 70 mm，纸箱内尺寸为 235 mm × 155 mm × 110 mm。饮料盒放置顺序遵循图 10 中 A ~F 顺序。随着饮料盒依次放入箱中，箱子结构特征发生变化，采用基于深度信息的物体定位方法进行定位。在饮料盒结构特征未发生变化情况下，使用 Gen6D 物体位姿估计方法进行定位。

首先进行视觉模式下约束放置实验。由于饮料盒材料具有弹性，接触纸箱底部时会发生形变，从而导致位置偏移，进而影响后续物体放置精度。在饮料盒进入纸箱后，视觉系统无法继续进行位姿校正。针对这一问题，采用柔顺控制策略，通过在 X、Y 轴方向施加柔性推压来校正物体位置。在校正过程中，当力传感器读数达到设定阈值时，表明物体已到达边界，此时系统自动停止调整，从而确保物体能够准确放置，并且不会影响后续物体堆放精度。

实验分为 4 组，每组实验纸箱放置位置不同，且每组进行 10 次放置实验。图 12 为 1 组视觉模式的物体约束放置完整的实验流程。实验中，成功的约束放置定义为将 6 个饮料盒准确放置到指定位置。若实验过程中发生物体碰撞或掉落，则该组实验视为失败。实验结果见表 2。实验中约束放置平均成功率为 30%，表明基于视觉的约束放置成功率较低，未能满足实际应用需求。

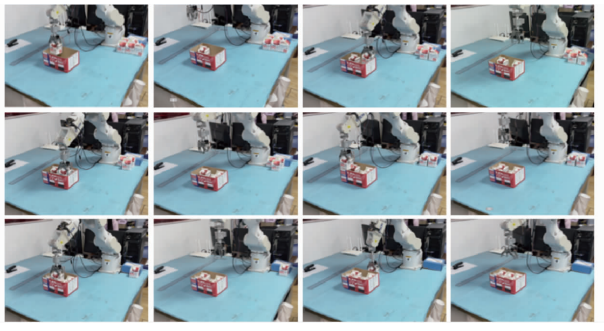


图 12　视觉模式的物体约束放置

Fig. 12　Object constrained placement in visual-mode

表 2　视觉模式约束放置实验结果

Tab. 2　Vision-mode constrained placement results

编号	成功次数	成功率/%	放置成功次数	放置成功率/%
1	3	30	35	58.3
2	2	20	26	43.3
3	4	40	36	60.0
4	3	30	31	51.6

失败的主要原因包括：①视觉定位误差大，定位不准，导致实验失败。②机械手爪设计缺陷，手爪抓取物体需要闭合角度，闭合角度小，容易导致舵机堵转进而烧毁。闭合角度大，被抓的饮料盒容易松动并倾斜，导致与纸箱边缘碰撞，从而导致实验放置失败。从表 2 可知，基于深度信息的定位方法在受限空间的作业任务下表现不佳。为提高约束放置实验成功率，需引入决策树算法进行实时碰撞预测和调整，补偿定位误差。

视觉-力觉模式下约束放置实验流程为：首先，机器人抓取饮料盒移动到纸箱目标区域上方，若视觉定位不准确，在向下放置过程中饮料盒底部会与纸箱上边缘发生碰撞。当力传感器监测到垂直方向力 F_z 发生突变时，判定碰撞发生。此时，机器人将饮料盒向上抬起，并触发决策树调整机制。将检测到的六维力数据传入决策树模型，模型预测出所需的调整偏移量(S_x,S_y)，并驱动机器人进行相应的移

动调整。

在物体向下放置过程中，$X+$、$X-$、$Y+$、$Y-$ 4 个方向上均发生过碰撞。图 13 为物体从到达目标点上方至实际放置过程中 $Y-$ 方向碰撞力与力矩变化。由图 13 可知，在 0 ~ 2 s 内，Z 方向接触力 F_z 稳定在约 1.2 N，而其余方向的力和力矩在 0 附近波动，这表明机器人已成功抓取饮料盒并平稳移动至目标区域上方；在 2 ~ 3 s 内，力传感器数据发生突变，表明饮料盒与纸箱发生碰撞。图 13 中，F_x 为正值，F_y 和 F_z 为负值，其中 F_x、F_y 和 F_z 分别突变为 0.8、-0.75、-0.3 N。根据此时力传感器，预测出的 Y 方向 (S_x, S_y) 为 (0, -8) mm；在 3 ~ 5 s 内，力与 0 ~ 2 s 内的力呈相似特征，表明在此阶段机器人抬起物体，并根据预测出的 (S_x, S_y) 调整机器人末端方向和步长，尝试重新向下放置物体；在 5 s 之后，六维力传感器读数稳定在 0 附近，机器人末端执行器不再受力，表明机械手爪已成功张开，饮料盒已顺利放置于纸箱内，本次放置任务顺利完成。

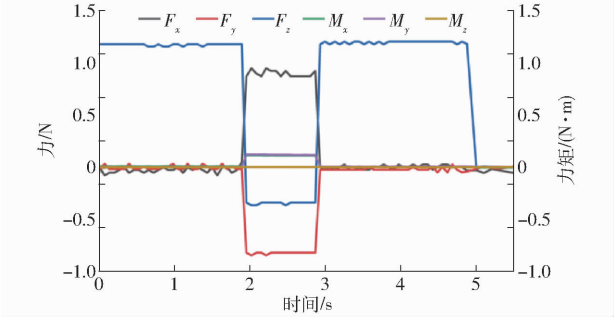


图 13 $Y-$ 方向碰撞力与力矩变化曲线

Fig. 13 $Y-$ axis collision force and torque variation

为准确评估决策树模型对物体约束放置实验结果的影响，实验分为 4 组，纸箱位置与视觉模式下的约束放置实验相同。视觉-力觉模式下约束放置实验结果见表 3。

表 3 视觉-力觉模式约束放置实验结果

Tab. 3 Vision - force-mode constrained placement results

编号	成功次数	成功率/%	放置成功次数	放置成功率/%	失败物体编号
1	8	80	53	88.3	D、C
2	9	90	56	93.3	C
3	8	80	54	90.0	D、D
4	7	70	47	78.3	B、C、C

位置 C 和 D 失败率较高，主要受限于空间狭窄及夹爪设计缺陷。如图 14 所示，位置 C 高频失效力觉曲线显示：1 ~ 2 s 发生首次碰撞，2 ~ 3 s 完成物体抓取；虽然决策树模型能正确预测调整方向和步长，但因夹爪设计缺陷导致物体倾斜，5 ~ 6 s 出现二次碰撞。F_z 始终不为 0，证实了饮料盒未能成功放入纸箱中，实验失败。但与基于视觉模式的约束放置实验相比，整体失败次数较少。

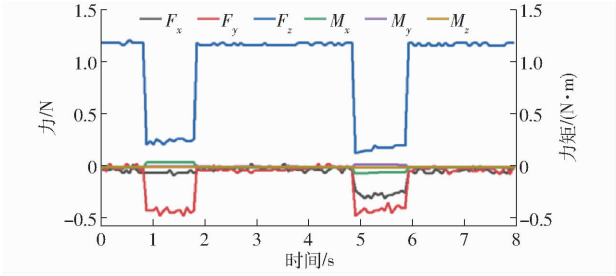


图 14 位置 C 失效力觉曲线

Fig. 14 Failure force curve at position C

由表 2 和表 3 可知，视觉模式约束放置实验平均成功率为 30%，放置成功率均值为 53%。然而，视觉-力觉模式的约束放置实验平均成功率提高至 80%，平均放置成功率提升至 87.5%。视觉-力觉模式约束放置实验成功率明显优于视觉模式约束放置实验成功率，验证了决策树模型有效修正视觉定位误差方面的能力。也表明了基于视觉与力觉的机器人交互系统在复杂环境下的高精度任务中表现出良好可靠性。通过视觉与力觉的协同感知与决策机制，显著增强了机器人与环境交互能力，提升了整体任务执行精度与稳定性。

5 结束语

提出了一种面向无模型物体的机器人视觉-力觉协同交互方法。该方法包括无模型物体视觉定位和基于力觉的物体交互策略 2 部分，通过三维重建、位姿估计算法及基于深度信息的物体定位方法实现无模型物体定位；根据机器人与环境间的接触力，结合决策树模型和柔顺控制算法动态调整机器人运动和操作。在搭建的实验平台上进行了抓取堆叠实验和约束放置实验。实验结果表明，提出的视觉-力觉协同交互方法能够实现对无模型物体定位，在复杂环境下执行高精度操作任务时展现出良好的可靠性，能显著提升机器人与环境之间的交互能力。

参 考 文 献

[1] 郧彦辉. 欧盟“工业 5.0”对我国智能制造发展的启示[J]. 中国外资, 2021(23): 46 - 47.
YUN Yanhui. The implications of “Industry 5.0” for China's intelligent manufacturing development[J]. Foreign Investment in China, 2021(23): 46 - 47. (in Chinese)

[2] SHAH S, BENNAMOUN M, BOUSSAID F, et al. Keypoints-based surface representation for 3D modeling and 3D object

recognition[J]. Pattern Recognition, 2017, 64: 29 - 38.
[3] HINTERSTOISSER S, HOLZER S, CAGNIART C, et al. Multimodal templates for real-time detection of texture-less objects in heavily cluttered scenes[C]// Proceedings of the 2011 International Conference on Computer Vision, 2011.
[4] HINTERSTOISSER S, LEPETIT V, ILIC S, et al. Model based training, detection and pose estimation of texture-less 3D objects in heavily cluttered scenes[C]// Computer Vision-ACCV, 2012.
[5] DROST B, ULRICH M, NAVAB N, et al. Model globally, match locally: efficient and robust 3D object recognition[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010.
[6] PENG S, ZHOU Xiaowei, LIU Yuan, et al. PVNet: pixel-wise voting network for 6DoF object pose estimation[C]// IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.
[7] CHARLES R Q, SU Hao, KAICHUN M, et al. PointNet: deep learning on point sets for 3D classification and segmentation[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2017.
[8] KEHL W, MANHARDT F, TOMBARI F, et al. SSD-6D: making RGB-based 3D detection and 6D pose estimation great again [C]// IEEE International Conference on Computer Vision, 2017.
[9] CACCAVALE F, CHIACCHIO P, MARINO A, et al. Six-DOF impedance control of dual-arm cooperative manipulators[J]. IEEE/ASME Transactions on Mechatronics, 2008, 13(5): 576 - 586.
[10] JUNG S, HSIA T C, BONITZ R G. Force tracking impedance control of robot manipulators under unknown environment[J]. IEEE Transactions on Control Systems Technology, 2004, 12(3): 474 - 483.
[11] 黄婷, 孙立宁, 王振华, 等. 基于被动柔顺的机器人抛磨力/位混合控制方法[J]. 机器人, 2017, 39(6): 776 - 785.
HUANG Ting, SUN Lining, WANG Zhenhua, et al. Hybrid force/position control method for robotic polishing based on passive compliance structure[J]. Robot, 2017, 39(6): 776 - 785. (in Chinese)
[12] 柯贤锋, 王军政, 何玉东, 等. 基于力反馈的液压足式机器人主/被动柔顺性控制[J]. 机械工程学报, 2017, 53(1): 13 - 20.
KE Xianfeng, WANG Junzhen, HE Yudong, et al. Active/passive compliance control for a hydraulic quadruped robot based on force feedback[J]. Journal of Mechanical Engineering, 2017, 53(1): 13 - 20. (in Chinese)
[13] 刘仁伟, 徐晓辉, 谢永权, 等. 基于机械臂辅助的卫星柔顺装配技术研究[J]. 机电工程, 2020, 37(5): 532 - 536.
LIU Renwei, XU Xiaohui, XIE Yongquan, et al. Compliant assembly technology of satellite assisted by robot arm[J]. Mechanical & Electrical Engineering Magazine, 2020, 37(5): 532 - 536. (in Chinese)
[14] FU Le, ZHAO Jie. Maxwell-model-based compliance control for human-robot friendly interaction[J]. IEEE Transactions on Cognitive and Developmental Systems, 2020, 13(1): 118 - 131.
[15] SARAVANAN A, JERALD J. Ontological model-based optimal determination of geometric tolerances in an assembly using the hybridized neural network and genetic algorithm[J]. Journal of Engineering Design, 2019, 30(4): 180 - 198.
[16] WANG Yuchen, WANG Peng, LIU Changchun, et al. Contact force/torque prediction and analysis model for large length-diameter ratio peg-in-hole assembly[C]// IEEE International Conference on Robotics and Biomimetics (ROBIO), 2018.
[17] EHSAN Z, WEN-FANG X. Robust AI-driven target-object-free hybrid vision/force control of industrial robotic systems[C]// IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2025.
[18] ZHANG Xianke, LI Zhongcan, ZHOU Yufei, et al. Hybrid vision/force cascaded model predictive control for automatic docking tasks[C]// IEEE Transactions on Industrial Electronics, 2025.
[19] 李道亮,赵晔,杜壮壮.农业领域多模态融合技术方法与应用研究进展[J]. 农业机械学报, 2025,56(1):1 - 15.
LI Daoliang, ZHAO Ye, DU Zhuangzhuang. Advances in multi-modal fusion techniques and applications in agricultural field [J]. Transactions of the Chinese Society for Agricultural Machinery,2025,56(1):1 - 15. (in Chinese)
[20] ÖZYESIL O, VORONINSKI V, BASRI R, et al. A survey of structure from motion[J]. Acta Numerica, 2017, 26: 305 - 364.
[21] HUANG Hongbo, YAN Xiaoxu, ZHENG Yaolin, et al. Multi-view stereo algorithms based on deep learning: a survey[J]. Multimedia Tools and Applications, 2024, 84: 2877 - 2908.
[22] LIU Yuan, WEN Yilin, PENG S D, et al. Gen6D: generalizable model-free 6-DoF object pose estimation from RGB images [C]// 17th European Conference on Computer Vision (ECCV), 2022.
[23] 黄玲涛, 王彬, 倪水, 等. 基于力传感器重力补偿的机器人柔顺控制研究[J]. 农业机械学报, 2020, 51(3): 386 - 393.
HUANG Lingtao, WANG Bin, NI Shui, et al. Robotic compliant control based on force sensor gravity compensation[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(3): 386 - 393. (in Chinese)
[24] 黄玲涛, 王彬, 倪涛, 等. 基于 Kinect 的机器人抓取系统研究[J]. 农业机械学报, 2019, 50(1): 390 - 399.
HUANG Lingtao, WANG Bin, NI Tao, et al. Research on robot grasping system based on kinect camera[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(1): 390 - 399. (in Chinese)
[25] 黄玲涛, 孔紫静, 杨帆, 等. 基于深度归一化的任意交互物体检测方法研究[J]. 农业机械学报, 2024, 55(8): 428 - 436.
HUANG Lingtao, KONG Zijing, YANG Fan, et al. Arbitrary interactive object detection method based on deep normalization [J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(8): 428 - 436. (in Chinese)