



基于6D位姿识别面向任意物体的智能人机协同递送

张树忠 朱祺 张弓 陈旭飞 杨根 吴月玉 齐春雨 邸思

Intelligent human-robot collaborative handover system for arbitrary objects based on 6D pose recognition

ZHANG Shuzhong, ZHU Qi, ZHANG Gong, CHEN Xufei, YANG Gen, WU Yueyu, QI Chunyu, DI Si

引用本文:

张树忠, 朱祺, 张弓, 陈旭飞, 杨根, 吴月玉, 齐春雨, 邸思. 基于6D位姿识别面向任意物体的智能人机协同递送[J]. *工程科学学报*, 2024, 46(1): 148–156. doi: 10.13374/j.issn2095-9389.2022.12.03.001

ZHANG Shuzhong, ZHU Qi, ZHANG Gong, CHEN Xufei, YANG Gen, WU Yueyu, QI Chunyu, DI Si. Intelligent human - robot collaborative handover system for arbitrary objects based on 6D pose recognition[J]. *Chinese Journal of Engineering*, 2024, 46(1): 148–156. doi: 10.13374/j.issn2095-9389.2022.12.03.001

在线阅读 View online: <https://doi.org/10.13374/j.issn2095-9389.2022.12.03.001>

您可能感兴趣的其他文章

Articles you may be interested in

基于深度循环神经网络的协作机器人动力学误差补偿

Error compensation of collaborative robot dynamics based on deep recurrent neural network

工程科学学报. 2021, 43(7): 995 <https://doi.org/10.13374/j.issn2095-9389.2020.04.30.003>

巡线机器人延迟容忍传感器网络数据传输策略

Date delivery scheme of delay-tolerant mobile sensor networks for high-voltage power transmission line inspection robot

工程科学学报. 2018, 40(11): 1412 <https://doi.org/10.13374/j.issn2095-9389.2018.11.015>

基于深度学习的人体低氧状态识别

Recognition of human hypoxic state based on deep learning

工程科学学报. 2019, 41(6): 817 <https://doi.org/10.13374/j.issn2095-9389.2019.06.014>

基于BP神经网络的机器人波动摩擦力矩修正方法

Wave friction correction method for a robot based on BP neural network

工程科学学报. 2019, 41(8): 1085 <https://doi.org/10.13374/j.issn2095-9389.2019.08.014>

基于群体智能优化的MKL-SVM算法及肺结节识别

MKL-SVM algorithm for pulmonary nodule recognition based on swarm intelligence optimization

工程科学学报. 2021, 43(9): 1157 <https://doi.org/10.13374/j.issn2095-9389.2021.01.14.004>

基于卷积神经网络的反无人机系统声音识别方法

Sound recognition method of an anti-UAV system based on a convolutional neural network

工程科学学报. 2020, 42(11): 1516 <https://doi.org/10.13374/j.issn2095-9389.2020.06.30.008>

基于 6D 位姿识别面向任意物体的智能人-机协同递送

张树忠¹⁾, 朱 祺^{1,2)}, 张 弓^{2,3)}✉, 陈旭飞^{1,2)}, 杨 根²⁾, 吴月玉²⁾, 齐春雨^{1,2)},
邱 思^{2,3)}

1) 福建理工大学福建省智能加工技术及装备重点实验室, 福州 350108 2) 广州先进技术研究所, 广州 511458 3) 中国科学院大学, 北京 100049

✉通信作者, E-mail: gong.zhang@giat.ac.cn

摘 要 在日常实践中存在大量人与人之间的多样性物体递送需求, 这可以依靠协作机器人来完成这些简单、耗时又耗力的任务。为此, 针对人-机协同递送过程中无法精确识别物体位姿导致难以准确抓取的问题, 引入基于 PnP 算法 (Perspective- n -Point) 的物体 6D 位姿识别网络, 实现待递送物体位姿的精确识别; 提出改进的被递送物体数据集制作方法, 实现面向任意物体的精准识别; 通过视觉系统标定、坐标转换以及抓取方案改进, 实现物体的精确位姿定位与准确抓取; 为验证所提出的人-机协同递送系统的有效性, 进行了基于 LineMod 数据集和自制数据集的人-机物体递送对比实验。结果表明, 面向自制数据集的物体递送提出的人-机递送系统平均误差距离为 1.97 cm, 递送平均成功率为 76%, 平均递送时间为 30 s; 如不考虑抓取姿势, 其递送成功率可达 89%; 具有较好的鲁棒性, 应用前景良好。

关键词 人-机协同; 物体递送; PnP 算法; 残差网络; 数据集

分类号 TP249

Intelligent human-robot collaborative handover system for arbitrary objects based on 6D pose recognition

ZHANG Shuzhong¹⁾, ZHU Qi^{1,2)}, ZHANG Gong^{2,3)}✉, CHEN Xufei^{1,2)}, YANG Gen²⁾, WU Yueyu²⁾, QI Chunyu^{1,2)}, QI Si^{2,3)}

1) Fujian Key Laboratory of Intelligent Machining Technology and Equipment, Fujian University of Technology, Fuzhou 350100, China

2) Guangzhou Institute of Advanced Technology, Guangzhou 511458, China

3) University of Chinese Academy of Sciences, Beijing 100049, China

✉Corresponding author, E-mail: gong.zhang@giat.ac.cn

ABSTRACT In daily practice, there are several instances of diverse object handover between humans. For example, in an automobile production line, workers need to pick up parts and deliver them to colleagues or acquire parts from them and put the parts in the appropriate position. Similarly, in households, children assist bedridden elderly people by passing them a cup of water, and in medical surgeries, assistants take over surgical tools used by doctors. These tasks require a considerable amount of time and manpower. In these scenarios, it is necessary to deliver the target object efficiently and quickly while prioritizing the safety of the object. Collaborative robots can serve as human colleagues to perform these simple, time-consuming, and laborious tasks. We expect humans and robots to hand over objects seamlessly in a natural and efficient way, just as humans naturally hand over objects to each other. This paper proposes a 6-dimensional (6D) pose recognition-based human-robot collaborative handover system to address the problem of inaccurate object

收稿日期: 2022-12-03

基金项目: 国家自然科学基金资助项目(62073092); 广东省自然科学基金资助项目(2021A1515012638); 福建省智能加工技术及装备重点实验室开放基金资助项目(KF-01-22005); 福建省级科技创新重点项目(2022G02007); 福建省 2022 年中央引导地方科技发展资金资助项目(2022L3014)

grasping caused by imprecise recognition of object poses during the human-robot collaborative handover process. The main contents are as follows: To solve the 6D pose recognition problem, a residual network (ResNet) is introduced to conduct semantic segmentation and key-point vector field prediction on the image, and the random sample consensus (RANSAC) voting is used to predict key-point coordinates. Further, an improved efficient perspective- n -point (EPnP) algorithm is used to predict the object pose, which can improve the accuracy. An improved dataset production method is proposed by analyzing the advantages and disadvantages of the LineMod dataset and based on the latest 3-dimensional (3D) reconstruction technology. To realize the accurate identification of daily objects, which can reduce the time required for dataset production. The transformation relationship (from the object to the camera and then to the robot base coordinate systems) is obtained through internal parameter calibration and hand-eye calibration methods of the camera. Thus, the pose of the target object in the robot base coordinate system is determined. Further, a grasping method for effective position and orientation calculation is proposed to realize precise object pose localization and accurate grasping. A handover experiment platform was set up to validate the effectiveness of the proposed human-robot collaborative handover system, with four volunteers conducting 80 handover experiments. The results showed that the average deviation distance of the proposed human-robot handover system is 1.97 cm, the average handover success rate is 76%, and the average handover time is 30 s, while the average handover success rate can reach 89% without considering the grasping posture. These results demonstrate that the proposed human-robot collaborative handover system is robust and can be applied to different scenarios and interactive objects with promising application prospects.

KEY WORDS human-robot collaboration; object handover; PnP algorithm; residual network; dataset

日常实践中存在大量人与人之间多样性的物体递送需求。例如: 在汽车生产线上, 工人需要提取零件将其交付给同事; 医疗手术中, 助手接过医生用完的手术工具; 家庭生活中, 子女帮助卧床的老人传递水杯。随着协作机器人的广泛应用, 这些看似简单却耗时耗力的任务都可让协作机器人作为人类的同事自然而有效地协同完成, 这就是人-机协同(Human-robot collaboration, HRC)^[1]。

现有人-机协同递送均需要专业设备配合^[2-3]。王宪伟等^[4]采用骨骼信息和 Red green blue-depth (RGB-D) 信息相结合的方式构建人机物体传递意图判别模型, 有效地区分了相同姿态表达的不同意图。Chan 等^[5]使用距离最小化方法实现了机器人自适应选择递送方向, 实现人-机递送, 但该方法未考虑递送物体的多样性。Liu 等^[6]使用点云的方式进行目标物体定位, 虽有较好的泛化能力, 但缺乏准确定位能力, 不适于精确抓取场景, 如无法选择抓取药瓶的瓶盖、瓶身, 也无法准确地抓取手术刀的刀柄。基于神经网络技术的 6D 位姿计算是解决此类问题的有效方法^[7]。例如, Xiang 等^[8]提出将 Pose convolutional neural network (PoseCNN) 网络用于计算目标物体的平移矩阵和旋转矩阵, 目标物体的平移矩阵通过相机的内外参数计算得到, 旋转矩阵通过卷积网络训练得出。Wu 等^[9]通过神经网络预测位姿的间接特征, 再计算得到物体的 6D 位姿。根据间接计算方式不同, 6D 位姿计算可以分为 N 点透视法 (Perspective- n -point, PnP)^[10]、特征描述方法^[11]和模型拟合方法^[12]。

因此, 本文提出一种基于 6D 位姿识别面向任意物体的智能人-机协同递送系统。通过识别空间中目标物体的 6D 位姿并选取相应的抓取姿势, 精确定位待递送物体位姿实现准确抓取; 基于点云重建系统制作自定义数据集, 实现了面向任意物体的精准识别, 扩展了递送系统使用场景; 进一步通过坐标转换, 将识别物体位姿从物体坐标系转换到机器人基坐标系中, 实现了从理论识别到实际抓取的过渡。

1 总体技术流程

针对递送过程中待抓取物体的定位问题, 本文采用间接法估计目标物体的 6D 位姿: 首先使用残差网络 (Residual network, ResNet) 对目标图片进行关键点向量预测以及语义分割^[13]; 其次根据随机抽样一致 (Random sample consensus, RANSAC) 投票机制对所有关键点进行评分, 再利用最远点采样法 (Farthest point sampling, FPS) 确定关键点的最终位置^[14-15]; 然后利用 PnP 方法对求得的关键点进行解算, 求得物体在空间中的位姿; 再通过标定手眼系统, 将识别位姿转换到世界坐标系中进行抓取; 最后通过人-机递送实验, 验证了该系统可行性和有效性。具体的技术流程如图 1 所示。

2 基于 ResNet 的物体 6D 位姿识别网络

2.1 基于 ResNet 的向量场预测和语义分割

理论上, 神经网络的层数越多, 其性能也会越好; 但在实际中, 过深的网络会产生退化问题, 性

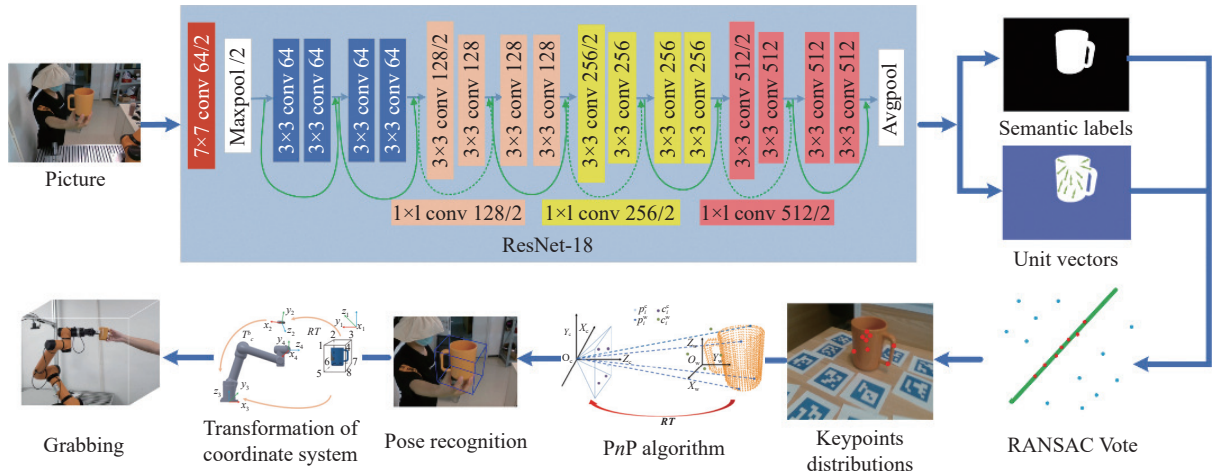


图 1 人-机协同递送系统技术流程

Fig.1 Technical flow of the human-robot collaborative handover system

能不增反减. He 等^[13]提出了 ResNet 结构来解决网络退化问题. 该结构在普通网络中新增一条分支, 将输入直接叠加到输出特征中.

以 ResNet-18 为主要结构对输入的 RGB 图像进行训练(图 1), 对其进行语义分割和关键点向量预测. 其结构主要包括全连接层、池化层和两个卷积层组成的残差块. 语义分割即将图片中的像素点进行分割, 从中挑选出目标物体的像素点; 关键点向量预测输出每个像素点到关键点的位置. 与传统方法中直接预测关键点的位置相比, 该方法使网络集中预测目标物体特征, 直接根据可见部分预估关键点位置, 避免关键点被遮挡时预测不准确.

2.2 RANSAC 投票

定义语义分割中识别的待抓取物体的某一像素点坐标为 p , 某一 2D 关键点 k 坐标为 x_k , 则网络中预测 p 点的向量为 $v_k(p)$, 表达式为:

$$v_k(p) = \frac{x_k - p}{\|x_k - p\|_2} \quad (1)$$

根据语义分割和关键点向量预测, 可在 RANSAC 投票中生成假设关键点. RANSAC 投票从一组包含了干扰点和关键点的数据集中, 通过迭代的方式估计关键点参数的数学模型. 由于 RANSAC 投票机制并不能每次都准确地确定模型, 需要进行多次迭代来提高正确率.

首先通过语义标签确认所有属于目标物体 O 的像素点, 找到随机两个像素对应向量的交点; 重复此过程 N 次, 得到一组假设关键点的集合 $\{h_{k,i} | i=1, 2, \dots, N\}$; 设目标物体 O 的像素点为集合 $h_{k,i}$, 计算 RANSAC 投票分数 $w_{k,i}$:

$$w_{k,i} = \sum_{p \in O} \Pi \left(\frac{(h_{k,i} - p)^T}{\|h_{k,i} - p\|_2} v_k(p) \geq \theta \right) \quad (2)$$

式中, Π 为指标函数; θ 为人工定义的一个阈值, 选取 0.99. 某点的投票分数越高代表其作为关键点的可能性越高, 生成关键点的空间概率分布. 假设关键点 x_k 的均值为 μ_k , 协方差为 \sum_k , 则:

$$\mu_k = \frac{\sum_{i=1}^N w_{k,i} h_{k,i}}{\sum_{i=1}^N w_{k,i}}, \quad \sum_k = \frac{\sum_{i=1}^N w_{k,i} (h_{k,i} - \mu_k)(h_{k,i} - \mu_k)^T}{\sum_{i=1}^N w_{k,i}} \quad (3)$$

为使该算法可以检测物体的不同角度, 可以使用最远点采样法将关键点均匀分布在目标物体 O 的表面. FPS 算法选取目标物体中心点为初始关键点, 然后寻找离该关键点最远的一个点, 将其添加至关键点集合 A 中, 重复多次, 直到该集合中的点达到 K 个, 在多次训练中 $K=8$ 时效果最佳.

2.3 PnP 求解

PnP 是求解 3D 点到 2D 点转换关系的方法, 即已知世界坐标系中 n 个关键点及其在 2D 图像坐标系中对应关键点坐标, 求得相机位姿的方法. 目前求解 PnP 问题主要有直接线性变换^[16]、非线性优化^[17]、P3P(Perspective-3-points)^[18]和 EPnP(Efficient perspective- n -points)方法^[19]. 直接线性变换只考虑了线性意义下的最优解没有考虑几何约束问题; 非线性优化将 PnP 问题构建为一个关于重投影误差的非线性最小二乘问题, 该方法准确率较高但计算量较大, 无法满足人-机协同递送中的实

时性要求; P3P 无法利用多于三组匹配点的信息, 容易造成缺解问题; EPnP 方法使用迭代的方式对相机位姿进行优化, 消除误差, 是目前的主流. 在 EPnP 方法的基础上, 引入空间概率分布, 采用莱文伯格-马夸特算法^[20](Levenberg-Marquardt algorithm)代替高斯-牛顿迭代法(Gauss-Newton iteration method)进行迭代计算, 优化了计算过程, 可增加位姿识别的准确性^[20-21].

PnP 算法如图 2 所示(以马克杯为例), 首先定义世界坐标系中的 n 个 3D 点 $\{p_i^w | i=1, 2, \dots, n\}$ 及其对应的图像坐标系中的 2D 关键点 $\{p_i^c | i=1, 2, \dots, n\}$. 世界坐标系和图像坐标系的关键点可以分别使用 4 个控制点 $\{c_j^w | j=1, 2, 3, 4\}$ 和 $\{c_j^c | j=1, 2, 3, 4\}$ 的权重归一化来表示:

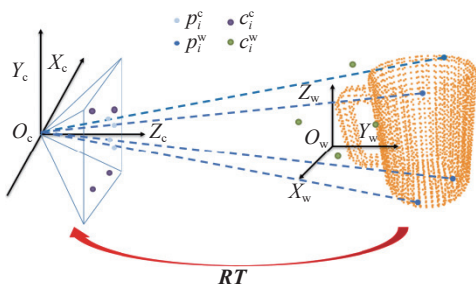


图 2 PnP 算法示意图

Fig.2 Schematic diagram of the PnP algorithm

$$p_i^w = \sum_{j=1}^4 \alpha_{ij} c_j^w; \quad p_i^c = \sum_{j=1}^4 \alpha_{ij} c_j^c; \quad \sum_{j=1}^4 \alpha_{ij} = 1 \quad (4)$$

式中, α_{ij} 为均质重心坐标(Homogeneous barycentric coordinates), 其世界坐标系和相机坐标系中的数值一致; w 为世界坐标系; c 为相机坐标系; i 和 j 为计数点. 原本的 3D-2D 问题转换为 3D-3D 问题, 即求解两组空间三维点之间的位姿变换. 2D 图像关键点可表示为:

$$\mathbf{K} p_i^c = \mathbf{K} \sum_{j=1}^4 \alpha_{ij} c_j^c = \begin{bmatrix} f_u & 0 & u_c \\ 0 & f_v & v_c \\ 0 & 0 & 1 \end{bmatrix} \sum_{j=1}^4 \alpha_{ij} \begin{bmatrix} x_j^c \\ y_j^c \\ z_j^c \end{bmatrix} \quad (5)$$

式中, \mathbf{K} 为相机内参矩阵; f_u 和 f_v 为相机内参矩阵中的焦距参数; u_c 和 v_c 为相机坐标系的中心坐标位置; x_j , y_j 和 z_j 为空间中第 j 个 3D 点的三维坐标. 由式(4)和(5)联立可得:

$$\begin{aligned} \sum_{j=1}^4 \alpha_{ij} f_u x_j^c + \alpha_j^i (u_0 - u_i) z_j^c &= 0; \\ \sum_{j=1}^4 \alpha_{ij} f_v y_j^c + \alpha_j^i (u_0 - u_i) z_j^c &= 0 \end{aligned} \quad (6)$$

将 n 个关键点串联, 可得:

$$\mathbf{M}x = 0 \quad (7)$$

式中, \mathbf{M} 为联立矩阵; x 为控制点坐标.

控制点在摄像头坐标系中的坐标为 $x = [c_1^{cT}, c_2^{cT}, c_3^{cT}, c_4^{cT}]^T$, 即:

$$x = \sum_{i=1}^N \beta_i v_i \quad (8)$$

式中, c_1^{cT} , c_2^{cT} , c_3^{cT} 和 c_4^{cT} 表示相机坐标系中的点; T 为转置; v_i 是 \mathbf{M} 中 N 个零特征值所对应的特征向量; β_i 为求解量. 原算法在计算出 β_i 的初始值后可以用高斯-牛顿迭代法对其进行迭代求解, 推导旋转矩阵 \mathbf{R} 和平移矩阵 \mathbf{T} . 基于 RANSAC 的投票得到关键点 x_k 的均值 μ_k 和协方差为 Σ_k . 通过最小化马哈诺比斯距离计算其 6D 位姿矩阵:

$$\text{Min}_{\mathbf{R}, \mathbf{T}} \sum_{k=1}^K \left((\tilde{x}_k - \mu_k)^T \Sigma_k^{-1} (\tilde{x}_k - \mu_k) \right); \quad \tilde{x}_k = \pi(\mathbf{R}X_k + \mathbf{T}) \quad (9)$$

式中, X_k 为关键点在世界坐标系中的三维坐标; \tilde{x} 为关键点在图像坐标系中的二维坐标; π 为透视投影函数; 旋转矩阵 \mathbf{R} 和平移矩阵 \mathbf{T} 的初始值由 EPnP 算法得出. 现使用莱文伯格-马夸特算法求解.

鉴于摄像头为固定安装, 则将求得的 \mathbf{RT} 齐次矩阵求逆后得到目标物体坐标系在摄像头坐标系中的位姿 rt .

3 改进的物体数据集制作

常用的 6D 位姿识别数据集有 Yale-CMU-Berkeley (YCB)^[22]、Texture-Less (T-LESS)^[23] 和 Line-Mod^[24], 但仅限于某几样物体, 不能完全覆盖日常使用场景. 而且由于人-机递送中所涉及的物体形状不规则、光学特性不统一以及相机存在局限性等因素, 基于 Augmented reality university of Cordoba (ArUco) 标定板^[25] 的三维重建方法效果不佳.

为此, 本文提出一种改进的数据集制作方法, 即以实际物体为参考进行 3D 建模并将其转换为点云模型, 然后将物体原始点云模型与第一帧 RGB 图像和深度图像构建的残缺点云重合定位, 使完整、准确的点云图像代替原本三维重建中效果不佳的点云模型, 如图 3 所示.

由于手动调整位姿矩阵无法使其完全重合, 引入了迭代最近点算法(Iterative closest point, ICP)进行点云匹配^[26]. 该算法通过迭代对应点搜寻和最小化点对整体距离以估计变换矩阵, 无需提取特征点且精度较高, 但容易陷入局部极小值. 因此

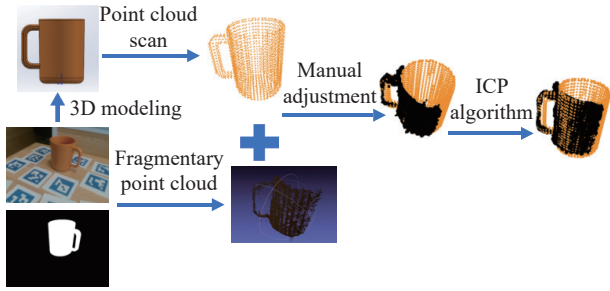


图 3 匹配点云模型

Fig.3 Point cloud matching model

在重合定位过程中,先手动调整使完整点云与残缺点云大部分重合,再进行 ICP 算法配准,以避免陷入局部最优.配准点云后,通过 ArUco 标定板输出语义分割图像和标准位姿,从而为神经网络训练提供数据.

与手动标注相比,ICP 算法简化了数据集的制作过程,可面向各种场景.图 4 为自制数据集的训练结果(为增加可靠性均进行了两次训练).其中虚线部分为第一次数据,实线部分为第二次数据.从图中可以看出,训练损失和投票训练损失在第 30000 次迭代时接近最小值;测试损失和投票测试损失在第 80 次迭代时接近最小值;ADD (Average 3D distance) 矩阵(预测位姿和真实位姿之间的平均距离)和 2D 投影矩阵(预测位姿和真实位姿的 3D 模型点投影之间的平均距离)来衡量预测位姿的准确性,且二者都在第 80 次迭代时接近最大值 1.

4 视觉系统标定及坐标转换

标定在视觉系统中是不可或缺的,其精度决

定递送的准确性.本文从相机成像原理分析开始,对递送系统进行相机内参及机器人手眼标定,将在神经网络中识别的物体位姿转换为世界坐标系中的位姿,以进行人-机协同物体递送实验,如图 5 所示.

4.1 手眼标定

为求解机器人基坐标系与相机坐标系之间的转换关系,采用眼在手外(Eye on hand)的形式^[27],即相机固定在机器人以外的基座上,不随机器人一起运动.

相机坐标系为 O_c , 标定板坐标系为 O_w , 机器人末端坐标系为 O_e , 机器人基坐标系为 O_b . 以及标定板坐标系到相机坐标系的位姿变换为 T_w^c , 相机坐标系到机器人基坐标系的位姿变换为 T_c^b , 机器人基坐标系到机器人末端坐标系的位姿变换为 T_b^e . 在标定过程中,将标定板固定在机器人末端,使其相对位姿保持不变.在机器人运动到某一位姿时,标定板上的一点在标定板坐标系 O_w 中的坐标为 P_1 , 其在机器人末端坐标系 O_e 中的坐标为 P_3 , 因而初始和变换位置后的相对关系为:

$$T_b^e T_c^b T_w^c P_1 = P_3; T_b^{e'} T_c^{b'} T_w^{c'} P_1 = P_3 \quad (10)$$

对上式进行联立转换,可得:

$$T_b^{e'-1} T_c^{b'} T_c^b T_w^c = T_c^{b'} T_w^{c'-1} T_w^c \quad (11)$$

通过多次变换机器人末端位姿,得到目标矩阵 T_b^e 的值.本文利用 ROS(Robot operating system) 中的 ros_aruco 功能包,获取标定板坐标系到相机坐标系的转换关系 T_w^c ; 通过遨博机器人 SDK(Software development kit) 获取机器人基坐标系到机器

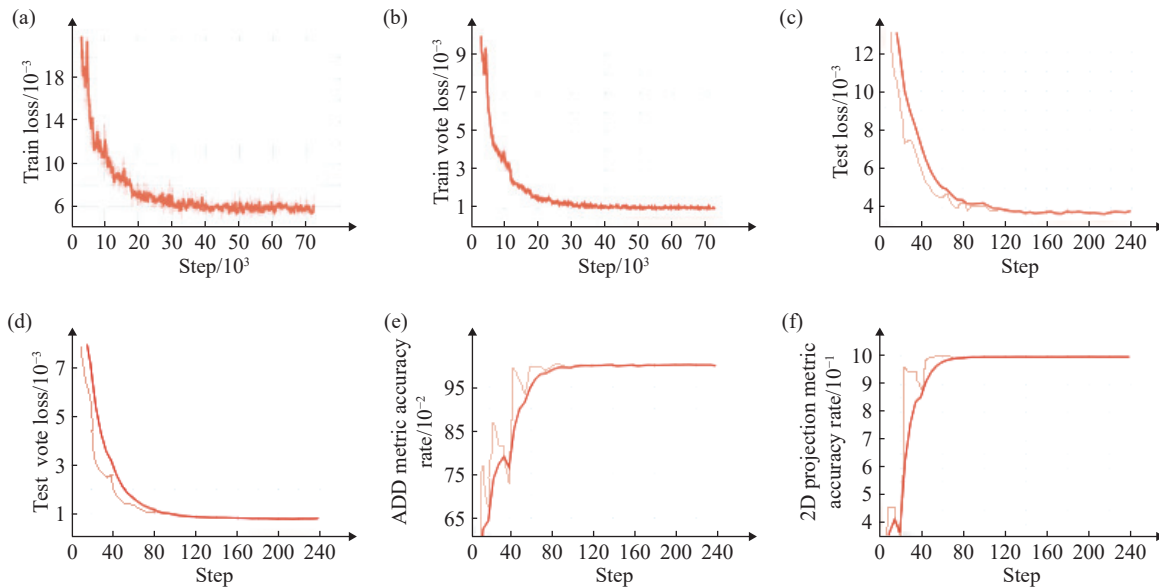


图 4 自制数据集训练结果. (a) 训练损失; (b) 投票训练损失; (c) 测试损失; (d) 投票测试损失; (e) ADD 矩阵; (f) 2D 投影矩阵

Fig.4 Training results of the self-made dataset: (a) train loss; (b) train vote loss; (c) test loss; (d) test vote loss; (e) ADD metric; (f) 2D projection metric

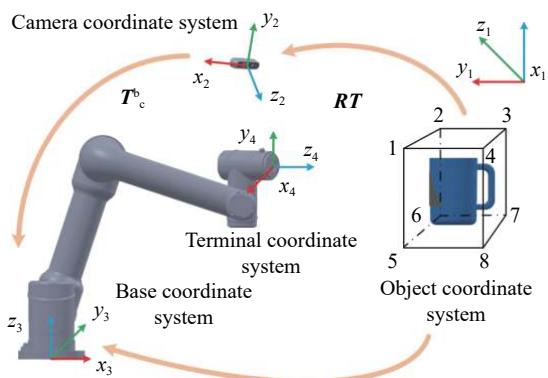


图5 物体坐标系到基坐标系转换过程

Fig.5 Transformation process of the object coordinate system to base coordinate system

人末端坐标系的位姿变换 T_b^e 代入 Tsai-Lenz 算法^[28] 中, 求解相机坐标系到机器人基坐标系的位姿变换 T_c^b , 其结果如式 (12) 所示:

$$T_c^b = \begin{bmatrix} -0.0389 & 0.5600 & -0.8276 & 0.4751 \\ 0.9992 & 0.0143 & -0.0373 & -0.4898 \\ -0.0090 & -0.8284 & -0.5601 & 0.9858 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (12)$$

4.2 坐标转换及改进的抓取方案

针对递送对象的不同, 为选取合适的抓取位姿, 在物体坐标系中选取可以完全覆盖目标物体的正方体结构, 其八个角点 $\{C_i^0 | i=1, 2, \dots, 8\}$ 即为目标点, 如图 5 所示。

基于前文所用 PnP 算法求得的目标物体坐标系与相机坐标系的转换矩阵 rt , 结合手眼标定得到的相机坐标系与机器人基坐标系的转换矩阵 T_c^b , 可求出机器人基坐标系中的角点坐标 $\{C_i^R | i=1, 2, \dots, 8\}$:

$$C_i^R = T_c^b \cdot rt \cdot C_i^0 \quad (13)$$

为增加递送系统的鲁棒性, 将抓取时的位置选取和姿势选取分开计算, 以确保即使抓取姿势有误, 也不影响最终效果。如图 5 所示, 假定最佳抓取点为水杯中心位置, 且平行于轴 $\vec{14}$ 的方向为最佳抓取姿势。则计算点 1 和 7 的坐标的平均值为抓取位置点; 选取轴 $\vec{67}$ 为 z 轴正方向, 轴 $\vec{62}$ 为 y 轴正方向, 轴 $\vec{63}$ 为 x 轴正方向, 以组成抓取坐标系。通过计算逆解, 使机器人末端原点到达抓取点, 且机器人末端坐标系与抓取坐标系平行, 即可完成标准位姿抓取。

5 人-机递送实验

5.1 实验平台及流程

人-机协同递送实验平台主要包括: Intel Real-

Sense D435i 摄像头, 其可作为人-机协同系统的“眼睛”, 负责拍摄实时图像并将其传递给工作站处理; 搭载了 NVIDIA 3090 GPU 的高性能工作站, 可作为系统中枢快速识别目标物体的位姿, 同时负责数据后续处理和传输; 遨博 (AUBO) i5 协作机器人, 负责接受物体位姿数据, 并进行递送操作, 如图 6 所示。

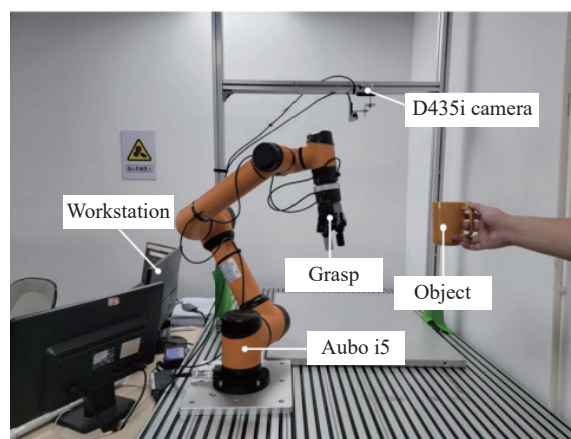


图6 人-机协同递送实验平台

Fig.6 Human-robot collaborative handover experimental platform

人-机协同物体递送操作的主要流程为: (1) 人类手持目标物体出现在摄像机和机器人的协同工作空间; (2) 待人类姿势稳定后, 拍摄照片并将其输入至神经网络以进行位姿识别; (3) 将识别后的位姿转换到机器人基坐标系, 计算逆解后传输给机器人; (4) 机器人运行到目标物体位置附近从人类手中抓取物体, 然后将其放到指定位置。

5.2 实验结果及分析

鉴于人-机协同物体递送中的诸多情况, 如人类身高的多样性会导致物体在空间中位置和姿态不同以及人类服装颜色对识别的干扰, 选取 4 位体态和穿着不一的志愿者, 每位均以不同的姿态、手持 LineMod 数据集的宠物猫物体, 各自进行 20 次递送实验。其递送过程包括: 识别、待抓取、已抓取、释放和放置; 物体位姿包括: 朝右上方、正前方和左下方, 如图 7 所示。

同时, 为验证自制数据集的可靠性, 以及人-机协同物体递送系统的鲁棒性, 也针对自制数据集中的马克杯物体, 进行了与 LineMod 数据集中物体相同模式的递送实验, 共 80 次, 如图 8 所示。

为客观量化抓取标准, 除误差距离外, 还设定了位置和姿势两种衡量标准。其中, 若机器人运动到最终位置后, 关闭夹爪时能保持抓紧物体而不掉落, 即位置达标; 若机器人抓取物体时, 机器人

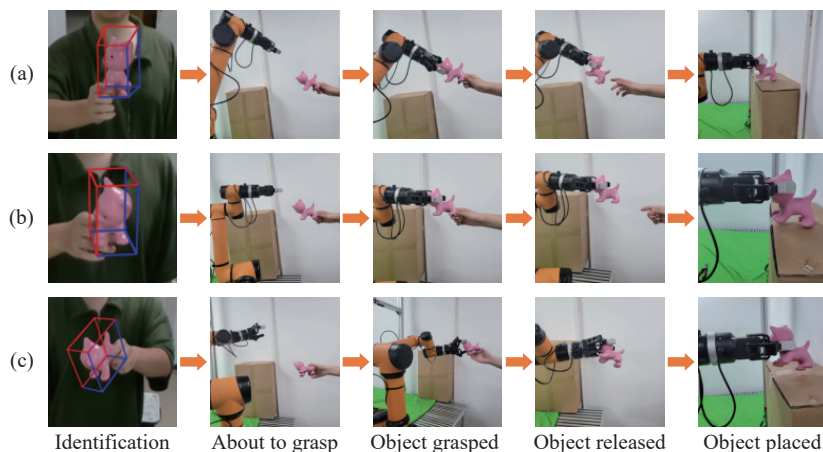


图 7 不同物体位姿情况下的递送流程(LineMod). (a) 朝右上方; (b) 正前方; (c) 左下方

Fig.7 Handover process (LineMod) at different object postures: (a) top right; (b) forward; (c) lower left

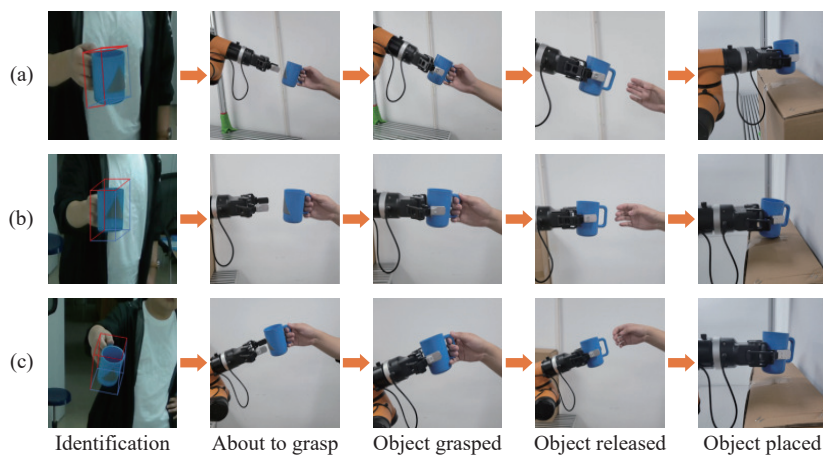


图 8 不同物体位姿情况下的递送流程(自制). (a) 朝右上方; (b) 正前方; (c) 左下方

Fig.8 Handover process at different object postures: (self-made dataset): (a) top right; (b) forward; (c) lower left

第六轴与目标物体保持平行, 则判断本次抓取姿势达标。

为保障志愿者安全, 实验中将机器人末端速度仅设置为 $0.5 \text{ m}\cdot\text{s}^{-1}$ 。结果表明: 神经网络平均识别时间为 4.3 s, 递送过程平均总时长约为 30 s。当递送物体的位置和姿势同时达标时, 则判定实验成功。面向 LineMod 数据集和自制数据集中物体的人-机协同递送实验, 其结果如表 1 所示。

从表中可以看出, 面向 LineMod 数据集的物体递送: 平均成功率为 88%; 平均误差距离为 0.925 cm; 在不考虑抓取姿势是否准确的情况下抓取成功率高达 95%, 基本达到了人-机协同递送要求。面向自制数据集的物体递送: 平均成功率为 76%、误差距离平均值为 1.97 cm; 在不考虑抓取姿势是否达标的情况下, 递送成功率可达 89%。相对于 LineMod 数据集, 其抓取成功率和平均误差距离虽然有所

表 1 人-机协同递送实验结果

Table 1 Experimental results of human-robot collaborative handover

Volunteers	Average deviation distance/cm		Ignoring pose qualification success rate/%		Considering pose qualification success rate/%	
	LineMod	Self-made	LineMod	Self-made	LineMod	Self-made
1	0.905	1.805	95	85	90	80
2	1.045	1.700	90	85	80	70
3	0.630	2.089	100	90	100	80
4	0.920	2.200	95	95	95	75
Average	0.925	1.970	95	89	88	76

降低,但自制数据集扩展了人-机协同递送系统的多样性,节省了面向新的抓取环境和对象的准备和计算时间,也是未来人-机协同递送系统的发展方向之一。

6 结论

本文提出一种基于 6D 位姿识别面向任意物体的智能人-机协同递送系统。引入基于 ResNet 算法的物体 6D 位姿识别网络,实现了待递送物体位姿的精确识别;提出改进的被递送物体数据集的制作方法,实现了面向任意物体的精准识别;开展从识别图像坐标系到机器人基坐标系的转换,实现了从理论识别到实际抓取的过渡;进行了面向 LineMod 数据集和自制数据集的物体递送对比实验,具有较好的实用性,应用前景广泛。鉴于自制数据集的成功率较低,难以进行随动性的人-机递送,未来的研究中,会进一步优化自制数据集的技术流程,提升位姿识别的准确性。同时增加人-机随动递送机制,以提升系统的鲁棒性。

参 考 文 献

- [1] Yang W, Paxton C, Mousavian A, et al. Reactive human-to-robot handovers of arbitrary objects // 2021 *IEEE International Conference on Robotics and Automation (ICRA)*. Xi'an, 2020: 3118
- [2] Pan M K X J, Knoop E, Bächer M, et al. Fast handovers with a robot character: Small sensorimotor delays improve perceived qualities // 2019 *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Macau, 2019: 6735
- [3] Kobayashi F, Okamoto K, Kojima F. Robot-human handover based on position and posture of human hand // 2014 *Joint 7th International Conference on Soft Computing and Intelligent Systems (SCIS) and 15th International Symposium on Advanced Intelligent Systems (ISIS)*. Kitakyushu, 2014: 918
- [4] Wang X W, Cong M, Liu D, et al. A human-to-robot handover method inspired by human behavior patterns. *Modul Mach Tool Autom Manuf Tech*, 2021(2): 1
(王宪伟, 丛明, 刘冬, 等. 人类行为模式启发的人机物体传递. *组合机床与自动化加工技术*, 2021(2): 1)
- [5] Chan W P, Pan M K X J, Croft E A, et al. An affordance and distance minimization based method for computing object orientations for robot human handovers. *Int J of Soc Robotics*, 2020, 12(1): 143
- [6] Liu H Y, Wang L H. Collision-free human-robot collaboration based on context awareness. *Robot Comput Integr Manuf*, 2021, 67: 101997
- [7] Kehl W, Manhardt F, Tombari F, et al. SSD-6D: Making RGB-based 3D detection and 6D pose estimation great again // 2017 *IEEE International Conference on Computer Vision (ICCV)*. Venice, 2017: 1530
- [8] Xiang Y, Schmidt T, Narayanan V, et al. PoseCNN: A convolutional neural network for 6D object pose estimation in cluttered scenes // 14th *Conference on Robotics -Science and Systems*. Online, 2018
- [9] Wu Y. *Research on 6D Pose Estimation Method of Transparent Object Based on Fusion of Texture and Depth Information* [Dissertation]. Guangzhou: Guangdong University of Technology, 2022
(吴勇. 纹理与深度信息融合的透明物体 6D 位姿估计方法研究 [学位论文]. 广州: 广东工业大学, 2022)
- [10] Zakharov S, Shugurov I, Ilic S. DPOD: 6D pose object detector and refiner // *IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, 2019: 1941
- [11] Wohlhart P, Lepetit V. Learning descriptors for object recognition and 3D pose estimation // *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, 2015: 3109
- [12] Sock J, Kasaei S H, Lopes L S, et al. Multi-view 6D object pose estimation and camera motion planning using RGBD images // 16th *IEEE International Conference on Computer Vision (ICCV)*. Venice, 2017: 2228
- [13] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition // 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, 2016: 770
- [14] Derpanis K G. Overview of the RANSAC Algorithm. *Image Rochester NY*, 2010, 4(1): 2
- [15] Dovrat O, Lang I, Avidan S. Learning to sample // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, 2019: 2760
- [16] Příbyl B, Zemčík P, Čadík M. Absolute pose estimation from line correspondences using direct linear transformation. *Comput Visi Image Underst*, 2017, 161: 130
- [17] Crosilla F, Maset E, Fusiello A. Procrustean photogrammetry: From exterior orientation to bundle adjustment // Cefalo R, Zieliński J, Barbarella M. *New Advanced GNSS and 3D Spatial Techniques*. Cham: Springer, 2018: 157
- [18] Persson M, Nordberg K. Lambda twist: An accurate fast robust perspective three point (p3p) solver // *Proceedings of the European Conference on Computer Vision (ECCV)*. Munich, 2018: 334
- [19] Gong X R, Lv Y W, Xu X P, et al. Pose estimation of omnidirectional camera with improved EPnP algorithm. *Sensors*, 2021, 21(12): 4008
- [20] Bilski J, Kowalczyk B, Marchlewska A, et al. Local levenberg-marquardt algorithm for learning feedforward neural networks. *J Artif Intell Soft Comput Res*, 2020, 10(4): 299
- [21] Qin Y W, Kavetski D, Kuczera G. A robust Gauss-Newton algorithm for the optimization of hydrological models: From standard Gauss-Newton to robust Gauss-Newton. *Water Resour*

- Res*, 2018, 54(11): 9655
- [22] Calli B, Singh A, Walsman A, et al. The YCB object and Model set: Towards common benchmarks for manipulation research // 2015 *International Conference on Advanced Robotics (ICAR)*. Istanbul, 2015: 510
- [23] Hodan T, Haluza P, Obdržálek Š, et al. T-LESS: An RGB-D dataset for 6D pose estimation of texture-less objects // 2017 *IEEE Winter Conference on Applications of Computer Vision (WACV)*. Santa Rosa, 2017: 880
- [24] Hinterstoisser S, Lepetit V, Ilic S, et al. Model based training, detection and pose estimation of texture-less 3d objects in heavily cluttered scenes // *Asian Conference on Computer Vision*. Berlin, 2012: 548
- [25] Fu Q, Chen X Y, Zheng Z L, et al. Research progress on visual perception system of bionic flapping-wing aerial vehicles. *Chin J Eng*, 2019, 41(12): 1512
(付强, 陈向阳, 郑子亮, 等. 仿生扑翼飞行器的视觉感知系统研究进展. 工程科学学报, 2019, 41(12): 1512)
- [26] Zhang J Y, Yao Y X, Deng B L. Fast and robust iterative closest point. *IEEE Trans Pattern Anal Mach Intell*, 2022, 44(7): 3450
- [27] Song C H, Zhang G; Liu X H, et al. *Machine Vision, Principles and Detailed Explanation of Classic Cases*. 1st Ed. Beijing: Chemical Industry Press, 2022
(宋春华, 张弓, 刘晓红, 等. 机器视觉, 原理与经典案例详解. 1 版. 北京: 化学工业出版社, 2022)
- [28] Tsai R Y, Lenz R K. A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. *IEEE Trans Robot Autom*, 1989, 5(3): 345