

DOI:10.3969/j.issn.1003-5060.2023.09.005

基于Mask R-CNN和关键点提取的抓取位姿估计方法

吴飞, 金圣洁, 林晓琛

(武汉理工大学机电工程学院, 湖北武汉 430070)

摘要:针对工业现场中目标工件抓取位姿任意、在空间中散乱堆叠遮挡的问题,文章提出一种基于掩膜区域卷积神经网络(mask region-based convolutional neural network, Mask R-CNN)和关键点提取的抓取位姿估计方法。对抓取环境中待抓取工件与不可抓取工件采用实例分割,并构建抓取目标表面点云;对目标表面点云进行三维尺度不变特征变换(3D scale invariant feature transform, 3D SIFT)关键点提取,对模板点云进行三维角点(3D Harris)关键点提取,以提取的关键点作为采样一致性初始配准(sample consensus initial alignment, SAC-IA)算法的初始值,减少点云配准的计算量,实现目标表面点云与参考模板点云粗配准,利用迭代最近点(iterative closest point, ICP)算法进行精确配准。以三通和弯管作为实验对象,通过与单特征算子提取方法对比分析,表明该方法能够进一步降低点云配准时间,提高配准精度。

关键词:位姿估计;实例分割;关键点提取;采样一致性初始配准(SAC-IA);迭代最近点(ICP)

中图分类号:TP391

文献标志码:A

文章编号:1003-5060(2023)09-1178-07

Grasping pose estimation method based on Mask R-CNN and key point extraction

WU Fei, JIN Shengjie, LIN Xiaochen

(School of Mechanical and Electrical Engineering, Wuhan University of Technology, Wuhan 430070, China)

Abstract: For arbitrary grasping pose of target workpieces scattered and occluded in industrial site, a grasping pose estimation method based on mask region-based convolutional neural network (Mask R-CNN) and key point extraction is proposed. The workpiece to be grasped and the workpiece not to be grasped are separated by instance segmentation in the grasping environment, and the point cloud of the grasping target surface is constructed. The key points, extracted from the target surface point cloud by 3D scale invariant feature transform (3D SIFT) and the template point cloud by 3D Harris, are used as the initial value of sample consensus initial alignment (SAC-IA) algorithm to reduce the computational burden of the point cloud registration and achieve the coarse registration of target surface point cloud and reference template point cloud. Then, iterative closest point (ICP) algorithm is used for accurate registration. Taking tee joint and elbow pipe as experimental objects, experiments were conducted by the comparison with single feature operator extraction method, which show that this method can further reduce the point cloud registration time and improve the registration accuracy.

Key words: pose estimation; instance segmentation; key point extraction; sample consensus initial alignment (SAC-IA); iterative closest point (ICP)

在工件堆叠摆放的场景中,目标工件在空间中存在不同位姿、相互遮挡,增加了机器人抓取工件的难度。传统三维设备扫描目标的整体环境,

生成的点云数量庞大,直接作为源点云与参考点云配准,会导致运算时间过长,难以满足机器人抓取所需的快速性。随着工业智能化的快速发展,

收稿日期:2021-12-09;修回日期:2022-03-09

基金项目:国家自然科学基金资助项目(51675393)

作者简介:吴飞(1973—),男,河南叶县人,博士,武汉理工大学教授,硕士生导师。

快速、准确地获取目标工件的抓取位姿,对工业机器人的进一步发展有重要意义。

文献[1]通过 ORB(oriented fast and rotated brief)算法和模板匹配进行目标识别,利用改进的迭代最近点(iterative closest point, ICP)算法进行点云配准与位姿估计,实现了放射源物体相对相机的位姿估计;文献[2]从图像中提取出仅有抓取目标的局部图像,运用卷积神经网络(convolutional neural network, CNN)进行目标识别分类,该方法具有识别精度高、抗干扰能力强的特点;文献[3]提出一种基于深度学习的机械臂最优抓取位置检测方法,该方法具有较强的泛化性和稳定性;文献[4]将抓取位姿转化为抓取位置回归与抓取角度分类的组合,实现了抓取位姿的单次预测,提升了检测速度;文献[5]提出一种基于级联卷积神经网络的平面抓取位姿快速检测方法,针对背景多样、光照不均、存在噪点的场景,能够快速准确地计算得到末端执行器位姿;文献[6]对图像的每个空间位置进行抓取检测,能够直接回归获得检测结

果,但并不适用于多对象的图像数据;文献[7]将卷积神经网络与支持向量机(support vector machine, SVM)等传统机器学习算法相结合,进一步提高了位姿估计的准确率;文献[8]提出一种基于图像掩膜估计抓取位姿的方法,训练神经网络时不需要进行真实的位姿标注,保证了识别的效率,同时降低了系统的复杂度。

通过将更快速的区域卷积神经网络(faster region-based convolutional neural network, Faster R-CNN)^[9]等目标检测网络引入到位姿估计问题中,对于平面简单放置的物体,实现了准确的位姿估计及抓取,但针对多个目标物体散乱堆叠的情况,难以达到抓取要求。基于上述问题以及工业现场中三通件和弯管件散乱堆叠、位姿任意的工况,本文提出一种基于掩膜区域卷积神经网络(mask region-based convolutional neural network, Mask R-CNN)^[10]和关键点提取混合的位姿估计方法,并通过仿真实验分析了该方法的性能。位姿估计系统流程如图 1 所示。

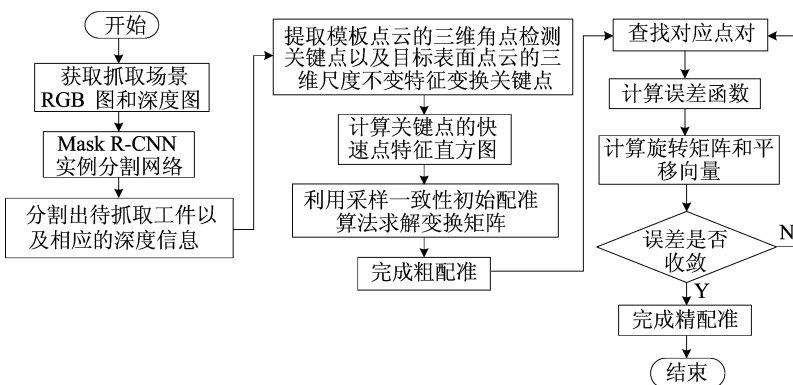


图 1 位姿估计系统流程图

1 Mask R-CNN 实例分割

1.1 制作数据集

利用 Realsense d435i 深度相机采集 200 张三通、弯管工件图像,图像像素大小为 848×480。图像采集环境如图 2 所示。

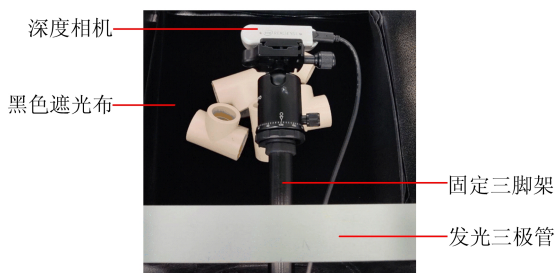


图 2 图像采集环境

深度相机通过固定三脚架安装在固定高度,采用黑色吸光布作为拍摄背景,通过调节曝光参数和发光二极管的亮度,可采集到工件与背景对比明显的数据集图像。

图像中包含不同倾斜角度的工件,同时采集部分含有不可抓取工件位姿的图像,作为负样本进行训练,以模拟真实的散乱堆叠工件场景,提高模型的鲁棒性,数据集如图 3 所示。在机器人抓取过程中,将所有识别出的工件抓取后,通过振动装置再次改变被识别为不可抓取工件的位姿,进行二次抓取。

由于采集的数据较少,直接进行训练容易出现过拟合现象,需要通过数据增广来提高实例分割模型的泛化能力。增广方式包括翻转、缩放、加

噪、模糊、加减曝光等方式,但并非所有的增广方式都适用于工件数据集,需要根据数据集的特征选择合适的数据增广方式。

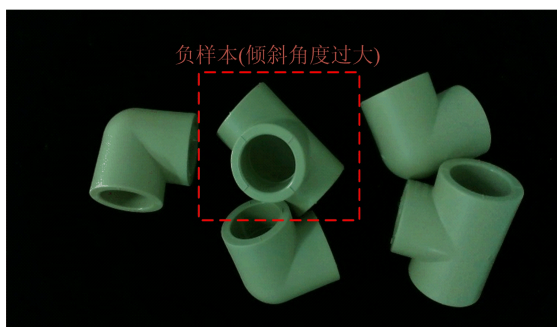


图 3 数据集

本次数据增广方式主要包括翻转、加噪、模糊及加减曝光,数据增广效果如图 4 所示。通过数据增广获得图像 3 000 张,其中 2 400 张作为训练集,600 张用于验证集。

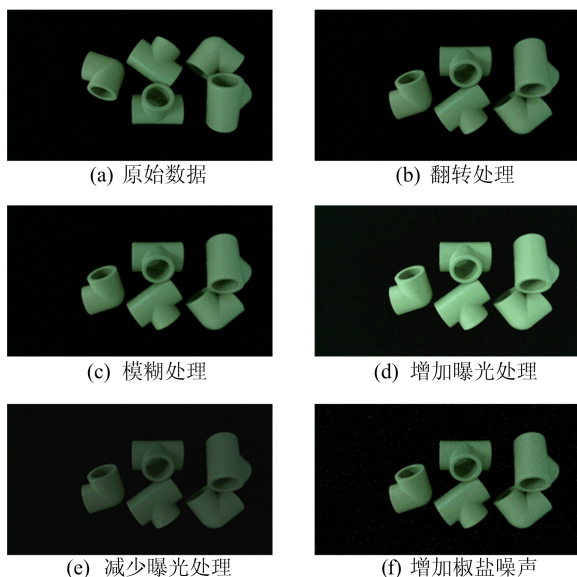


图 4 数据增广效果

1.2 Mask R-CNN 模型

Mask R-CNN 是由 Faster R-CNN 逐渐改进而来的多任务卷积网络模型,可以完成目标检测、目标分类、像素级别的目标分割等多种任务,Mask R-CNN 网络框架图如图 5 所示。

本文目标是实现工件分割与位姿估计,检测分割场景中含有许多同类的不同工件个体,无法采用语义分割模型。同时工件摆放位姿任意,存在堆叠遮挡的情况,难以从单一或者几个有限的角度分析出有意义的物体特征,也不可能穷举描述每一个特征,因此传统物体识别方法难以在该问题中发挥作用,故采用 Mask R-CNN 实例分割模型对目标工件

的不同个体进行分割。

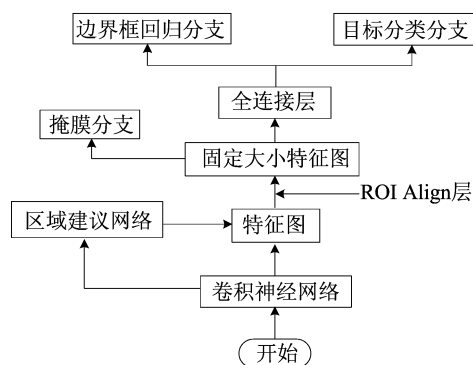


图 5 Mask R-CNN 网络框架图

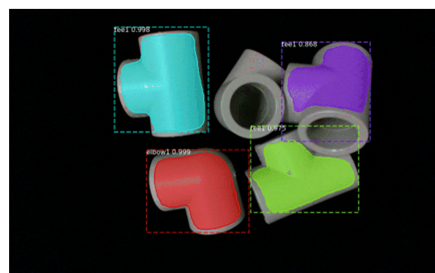
1.3 Mask R-CNN 训练与验证

基于迁移学习的方式,加快训练的过程,以 COCO(common objects in context)数据集的训练权重作为初始化权重,在 AMD 2600X @3.6 GHz,内存 32 GiB, GPU NVIDIA GTX-2070 的 Windows10 平台 TensorFlow1.8.0 环境下进行 Mask R-CNN 的训练。模型训练参数设置见表 1 所列。

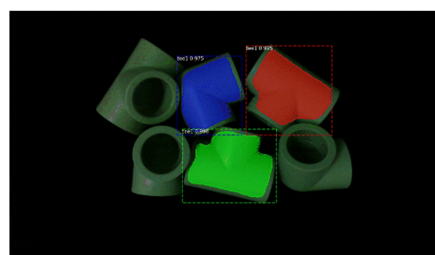
表 1 模型训练参数

参数名称	数值	说明
BACKBONE	ResNet50	特征提取
MOMENTUM	0.9	下降动量
LEARNING_RATE	0.001	学习率
EPOCH	30	遍历次数

输出掩膜结果如图 6 所示,Mask R-CNN 实例分割网络能够准确地识别出待抓取的目标工件以及位姿偏差过大的工件,同时能够识别出部分被遮挡的工件。



(a) 三通件部分遮挡



(b) 弯头件部分遮挡

图 6 Mask R-CNN 输出掩膜结果

1.4 表面点云的构建

通过Mask R-CNN对彩色图进行分割,获得的掩膜mask可对深度图进行分割并提取深度信息,如图7所示。

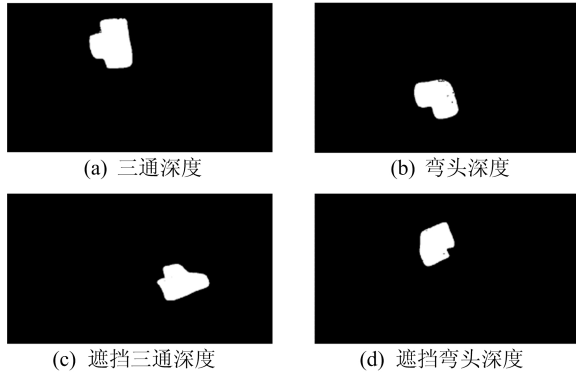


图7 分割工件深度

利用深度相机的内参可建立工件表面点云,点云中每个点的坐标 (X, Y, Z) 的计算公式如下:

$$\begin{cases} X = (u - u_0)Z / f_x, \\ Y = (v - v_0)Z / f_y, \\ Z = d/s \end{cases} \quad (1)$$

其中: u_0, v_0 为相机光圈 X, Y 方向的中心点位置; f_x, f_y 分别为 x, y 轴上的焦距; s 为深度图里的数据与实际距离的比例。

根据式(1)计算结果分割出待抓取工件表面点云,如图8所示。

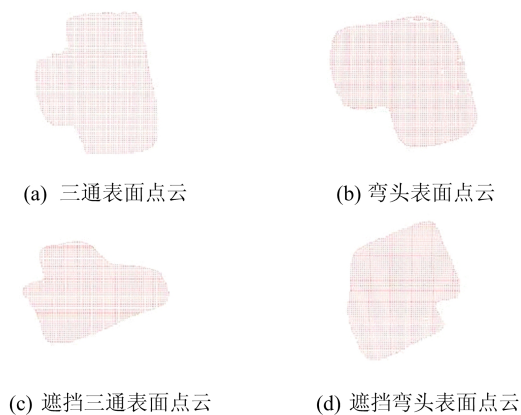


图8 分割工件点云图

2 点云配准方法

2.1 特征关键点对比

由于点云数据是三维坐标系中采样点的数据集,且实际获得的点云数据往往数量级较大,对不同坐标系下的所有点迭代计算刚性变换矩阵,不

仅需要较长的运算时间,也极易造成局部收敛。因此,可采用点云的关键点去描述一个点邻域内的点云,不但减少点云数量,还提高配准效率和配准精度。常用的点云关键点提取算子有三维尺度不变特征变换(3D scale invariant feature transform, 3D SIFT)算子^[11]、三维角点(3D Harris)检测算子^[12]等。

本文依次在体素下采样后三通工件的计算机辅助设计(computer aided design, CAD)模板点云、无遮挡工件表面点云以及有遮挡工件表面点云测试3D Harris、3D SIFT 2种算子,CAD模板点云数量为6 621个,无遮挡工件表面点云数量为3 033个,遮挡工件表面点云数量为2 849个,关键点提取结果对比如图9所示。

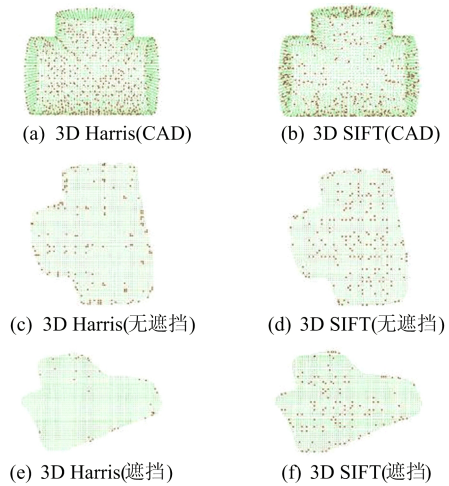


图9 关键点提取结果对比

根据上述关键点提取结果,针对CAD模板点云,3D Harris关键点分布较为均匀,能够描述整体点云的特征,符合关键点提取要求,3D SIFT关键点主要分布在工件边缘处,不利于后续与表面点云的配准。针对工件的表面点云,存在无遮挡表面点云和有遮挡情况下表面点云2种情况,如图9结果所示,3D Harris关键点相较于3D SIFT关键点主要集中在工件边缘处,且关键点数量相对较少,难以体现表面点云的整体特征。关键点提取数量见表2所列,关键点提取时间见表3所列,配准时间以关键点提取10次的平均时间为准。

由统计结果可知,3D Harris算子相比3D SIFT算子在关键点提取上所需时间更短,结合图9关键点提取结果可知,由于CAD模板点云数量较多,可选用3D Harris提取算法,减少计算时

间。针对分割工件表面点云,3D SIFT 关键点提取时间与 3D Harris 相差 1 s 左右,但关键点提取更均匀,故可选用 3D SIFT 提取算法。

表 2 关键点提取数量 单位:个

点云类型	3D Harris	3D SIFT
模板点云	580	605
无遮挡点云	143	269
有遮挡点云	31	153

表 3 关键点提取时间 单位:s

点云类型	3D Harris	3D SIFT
模板点云	3.029	8.584
无遮挡点云	0.482	1.921
有遮挡点云	0.348	1.308

2.2 初步位姿估计

本文采用采样一致性初始配准算法(sample consensus initial alignment, SAC-IA)^[13]对模板点云和目标工件表面点云进行初始位姿配准,该算法以点云的快速点特征直方图(fast point feature histogram, FPFH)作为输入,通过计算特征之间的对应关系,完成初始配准。算法主要流程如下所述。

设定最小距离阈值 d_s ,从模板点云 P 中选取 n 个采样点,每个采样点之间的距离需大于阈值 d_s ,确保采样点具有不同的 FPFH。同理,计算目标点云 Q 的 FPFH,查找与点云 P 中采样点具有相似 FPFH 的 1 个或多个点,将其作为模板点云 P 中采样点在目标点云 Q 中的对应点。计算对应点之间旋转、平移变换矩阵,求解对应点变换后的配准误差来判断当前配准变换的性能。采用 Huber 公式计算配准误差 E_s ,即

$$E_s = \sum_{i=1}^n H(l_i) \quad (2)$$

$$H(l_i) = \begin{cases} \frac{1}{2} l_i^2, & l_i \leq l_m; \\ \frac{1}{2} l_m (2 | l_i | - l_m), & l_i > l_m \end{cases} \quad (3)$$

其中: l_m 为预先设定的距离阈值; l_i 为第 i 组对应点变换后的距离差。重复上述操作直至误差函数的值最小,即可求得所需的最佳变换矩阵。

SAC-IA 求得的变换矩阵并不准确,因此只能用于粗配准,初始点云如图 10 所示,无关键点提取粗配准效果如图 11 所示。

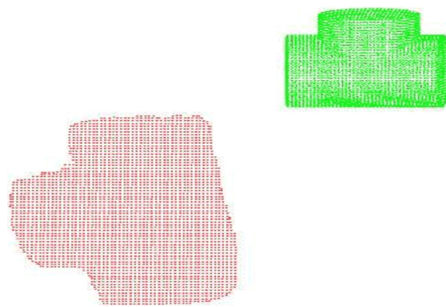


图 10 初始点云

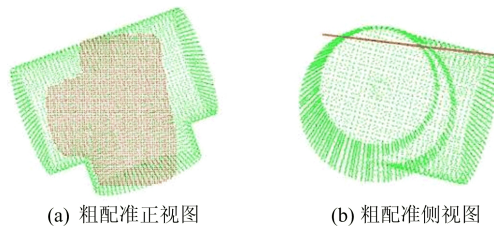


图 11 无关键点提取的 SAC-IA 粗配准效果

2.3 精配准位姿估计

通过粗配准两点云大致重合到一起,为 ICP 算法^[14]提供了良好的初始位姿。将粗配准后的两点云 P' 和 Q 作为 ICP 的初始点云集,算法实现的主要流程如下所述。

对于模板点云 P' 中的每个采样点矢量 p_i ,在目标工件表面点云 Q 中寻找距离最近点矢量 q_i ,确定初始最近对应点对 (p_i, q_i) 。计算对应点对之间旋转、平移变换矩阵 R, T ,使得对应点集之间的均方误差最小,均方误差计算公式如下:

$$d_k = \frac{1}{N} \sum_{i=1}^N \| q_i - (R p_i + T) \| \quad (4)$$

将求解得到的 R, T 变换矩阵作用于模板点云 P' ,得到变换后的点云 P'' 。设定阈值 $\epsilon = d_k - d_{k-1}$ 和最大迭代次数 N_{max} ,计算变换后的模板点云 P'' 和目标工件点云 Q 的均方误差,若误差小于阈值 ϵ 或者当前迭代次数超出最大迭代次数 N_{max} ,则停止计算;否则将粗配准获得的模板点云更新为 P'' ,继续重复上述步骤,直至满足收敛条件。

以 SAC-IA 的输出结果作为 ICP 精配准初始点云位姿,配准结果如图 12 所示。

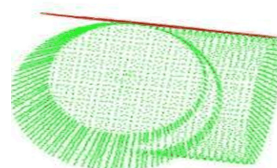


图 12 无关键点提取的 ICP 精配准效果

相较于图 11b 中粗配准效果,通过 ICP 精配准可以获得更为准确的位姿估计。

3 配准结果与分析

为了验证本文所提出方法的性能,对比分析了无关键点提取、3D Harris 关键点提取、3D SIFT 关键点提取、本文 3D Harris-3D SIFT 关键点提取混合的点云配准效率以及配准结果,点云实验的硬件环境为 Intel(R) Core(TM) i5-9400 CPU@2.90 GHz 处理器,16 GiB 内存。

以体素下采样后的三通件和弯管件点云作为算法对比的输入点云,无关键点提取精配准结果如图 11、图 12 所示,关键点提取算法的相关粗、精配准结果如图 13~图 18 所示。

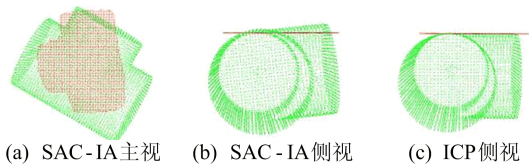


图 13 三通件 3D Harris 关键点提取结果

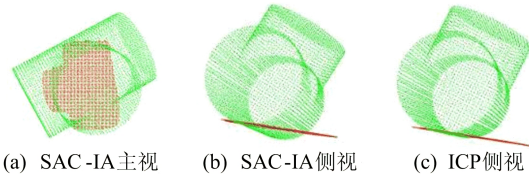


图 14 三通件 3D SIFT 关键点提取结果

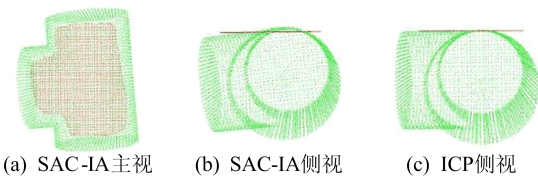


图 15 三通件 3D Harris-3D SIFT 关键点提取结果

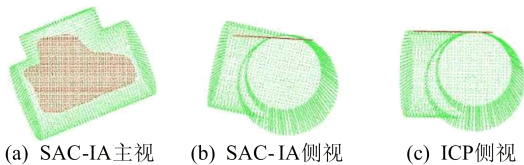


图 16 三通件遮挡点云 3D Harris-3D SIFT 关键点提取结果

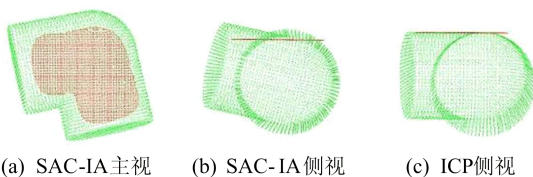


图 17 弯管件 3D Harris-3D SIFT 关键点提取结果

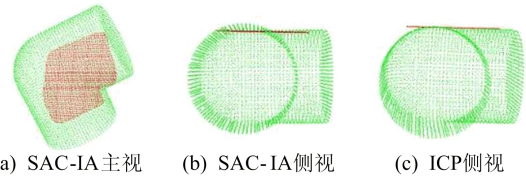


图 18 弯管件遮挡点云 3D Harris-3D SIFT 关键点提取结果

三通件配准数据统计结果见表 4 所列,配准对比分析结果见表 5 所列,配准时间以 10 次运行结果的平均值为准。由对比分析结果可知,无关键点提取输入的点云数量较大,配准时间长,不符合机器人抓取动作所要求的快速性;仅使用 3D Harris 关键点提取,针对模板点云,关键点提取较为均匀,而工件表面点云关键点提取数量少,粗配准效果差,但 3D Harris 关键点提取时间短,相较于无关键点提取,计算时间有明显的下降;仅使用 3D SIFT 关键点提取,针对模板点云,关键点主要集中在工件的边缘处,而工件表面点云提取较为均匀,导致配准结果出现在工件边缘处,位姿估计错误;而本文所提出的混合算法能够同时兼顾 2 种关键点提取算法的优点,点云特征明显,虽然比仅使用 3D Harris 关键点提取算法的配准时间略有上升,但点云的配准精度有显著提升,相较于仅使用 3D SIFT 关键点提取算法,配准时间更短,配准效果更好。

表 4 三通件配准点云数量 单位:个

对比项目	无遮挡	有遮挡
源点云	6 621	3 033
3D Harris 关键点	143	31
3D SIFT 关键点	269	153

表 5 三通件配准时间 单位:s

对比项目	无遮挡	有遮挡
无关键点配准	198.117	197.949
3D Harris 配准	19.897	18.672
3D SIFT 配准	34.161	26.498
3D Harris-3D SIFT 配准	22.773	19.056

4 结 论

针对工业现场中散乱工件的检测和位姿估计问题,本文提出了一种基于 Mask R-CNN 和关键点提取混合的位姿估计算法。通过 Mask R-CNN 能够准确地识别出待抓取的目标工件以及位姿偏差过大的不可抓取工件。通过对模板点云进行 3D Harris 关键点提取,目标工件表面点云

进行 3D SIFT 关键点提取,将模板点云与目标工件表面点云缩至特征点集,既保留了点云的主要特征,又能够提高点云配准效率。以三通和弯管工件为实验对象,实验对比分析了无关键点提取、3D Harris 关键点提取、3D SIFT 关键点提取以及本文关键点提取混合算法,结果表明,本文算法能够显著降低点云配准时间,提高配准精度,为后续的机器人抓取提供更加快速准确的位姿估计。

[参 考 文 献]

- [1] 曹文武. 基于 RGB-D 视觉识别的机器人抓取规划研究[D]. 哈尔滨:哈尔滨工业大学,2018.
- [2] 伍锡如,黄国明,孙立宁. 基于深度学习的工业分拣机器人快速视觉识别与定位算法[J]. 机器人,2016,38(6):711-719.
- [3] 杜学丹,蔡莹皓,鲁涛,等. 一种基于深度学习的机械臂抓取方法[J]. 机器人,2017,39(6):820-828,837.
- [4] 张云洲,李奇,曹赫,等. 基于多层级特征的机械臂单阶段抓取位姿检测[J]. 控制与决策,2021,36(8):1815-1824.
- [5] 夏晶,钱堃,马旭东,等. 基于级联卷积神经网络的机器人平面抓取位姿快速检测[J]. 机器人,2018,40(6):794-802.
- [6] REDMON J,ANGELOVA A. Real-time grasp detection using convolutional neural networks[J]. Proceedings IEEE International Conference on Robotics & Automation,2015:1316-1322.
- [7] ALEXANDRE L A. 3D object recognition using convolutional neural networks with transfer learning between input channels[C]//Proceedings of the 13th International Conference IAS-13. [S. l.]: Springer International Publishing,2016:889-898.
- [8] RAD M,LEPETIT V. BB8: A scalable, accurate, robust to partial occlusion method for predicting the 3D poses of challenging objects without using depth[C]//2017 IEEE International Conference on Computer Vision (ICCV). [S. l.]: IEEE,2017:3828-3836.
- [9] REN S,HE K,GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2017,39(6):1137-1149.
- [10] HE K,GKIOXARI G,P DOLLAR P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2017:2961-2969.
- [11] 俞浩,高飞. 基于特征向量的点云配准方法研究[J]. 合肥工业大学学报(自然科学版),2020,43(7):958-962.
- [12] SIPIRAN I,BUSTOS B. Harris 3D: a robust extension of the Harris operator for interest point detection on 3D meshes[J]. Visual Computer,2011,27(11):963-976.
- [13] 陈学伟,朱耀麟,武桐,等. 基于 SAC-IA 和改进 ICP 算法的点云配准技术[J]. 西安工程大学学报,2017,31(3):395-401.
- [14] 荆路,武斌,方锡禄. 基于 SIFT 特征点结合 ICP 的点云配准方法[J]. 激光与红外,2021,51(7):944-950.
- (责任编辑 胡亚敏)
-
- (上接第 1170 页)
- [3] CARMEL M. Soft-matter engineering for soft robotics[J]. Advanced Materials Technologies,2019,4(2):1800477.
- [4] CIANCHETTI M,LASCHI C,MENCIASSI A, et al. Biomedical applications of soft robotics[J]. Nature Reviews Materials,2018,3:143-153.
- [5] PENG Z,HUANG J. Soft rehabilitation and nursing-care robots: a review and future outlook[J]. Applied Sciences,2019,9(15):3102.
- [6] POLYGERIONS P,WANG Z,GALLOWAY K C, et al. Soft robotic glove for combined assistance and at-home rehabilitation[J]. Robotics and Autonomous Systems,2015,73:135-143.
- [7] SUN Z S,GUO Z H,TANG W. Design of wearable hand rehabilitation glove with soft hoop-reinforced pneumatic actuator[J]. Journal of Central South University,2019,26(1):106-119.
- [8] NING M,WANG K,LI M, et al. Design, analysis and experiment of finger soft actuator with nested structure for rehabilitation training[J]. Advances in Mechanical Engineering,2020,12(11):1-15.
- [9] HEUNG K,TONG R,LAU A, et al. Robotic glove with soft-elastic composite actuators for assisting activities of daily living[J]. Soft Robotics,2019,6(2):289-304.
- [10] WANG J B,FEI Y Q,WU P. Design, modeling, and testing of a soft pneumatic glove with segmented PneuNets bending actuators[J]. IEEE/ASME Transactions on Mechatronics,2019,24(3):990-1001.
- [11] 戴红. 人体运动学[M]. 北京:人民卫生出版社,2008:103-115.
- [12] 吴小瑜,周晓明. 人体损伤程度鉴定标准中手功能丧失的评定[C]//全国第十七届法医临床学学术研讨会论文集. 北京:中国法医学会,2014:628-629.
- [13] 管国阳,孟政委,谢立新,等. 聚氨酯橡胶超弹性本构模型拟合效果评估[J]. 力学季刊,2021,42(3):571-580.
- (责任编辑 张 镗)