

基于点云实例分割的鲁棒多目标位姿估计

刘耀华^{1,2,3,4*}, 马钺^{1,2,4}, 许敏^{1,2,4}¹中国科学院沈阳自动化研究所, 辽宁 沈阳 110016;²中国科学院机器人与智能制造创新研究院, 辽宁 沈阳 110169;³中国科学院大学, 北京 100049;⁴辽宁省智能检测与装备技术重点实验室, 辽宁 沈阳 100179

摘要 针对多目标位姿估计过程中点云局部特征存在类间错误匹配的问题,提出了基于点云实例分割的鲁棒多目标位姿估计算法。首先,基于密度聚类对场景点云进行分割得到点云簇,并用快速点特征直方图(FPFH)描述子对分割后的点云簇进行局部特征提取;然后利用随机森林算法对聚合后的点云簇的局部特征进行分类,完成点云实例分割;之后对于场景中每一个分割后的实例,采用近似近邻快速库(FLANN)匹配算法对场景实例和模型进行特征匹配,得到实例分割后的点在对类别模型上的匹配点,利用随机采样一致(RANSAC)算法以及最小二乘法计算初始位姿;最后经过点到平面迭代闭合点(ICP)算法得到每个实例的精确位姿。在 CV-Lab 3D 合成数据集以及 UWA 真实采集数据集下的实验结果表明,相比直接匹配模型和全部场景点的局部特征进行多目标位姿估计,所提算法能够有效提升局部特征匹配阶段的内点概率,从而提升复杂场景下位姿估计的鲁棒性和准确率,尤其适用于场景中具有多个实例的位姿估计应用。

关键词 图像处理;点云;实例分割;位姿估计;局部特征匹配;随机森林

中图分类号 TP391 文献标志码 A

DOI: 10.3788/LOP220586

Point-Cloud Instance Segmentation-Based Robust Multi-Target Pose Estimation

Liu Yaohua^{1,2,3,4*}, Ma Yue^{1,2,4}, Xu Min^{1,2,4}¹Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, Liaoning, China;²Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110169, Liaoning, China;³University of Chinese Academy of Sciences, Beijing 100049, China;⁴Key Laboratory on Intelligent Detection and Equipment Technology of Liaoning Province, Shenyang 100179, Liaoning, China

Abstract To address the problem of local features in point clouds being prone to mismatch between classes during multi-target pose estimation, a robust multi-target pose estimation algorithm based on point-cloud instance segmentation is proposed. First, point-cloud clusters are obtained by segmenting the scene point clouds based on density clustering, and the local feature of the point-cloud clusters are extracted using fast point feature histogram (FPFH) descriptor to describe the local geometry of the point clouds. Then, the random forest classifier is used to classify the aggregated local features of the point-cloud cluster, obtain the category to which the point-cloud cluster belongs, and completes the point-cloud instance segmentation. For each instance in the scene, the features of the scene instance and model are matched using the fast library for approximate nearest neighbors (FLANN) matching algorithm, and the matching points of the points after instance segmentation are obtained on the corresponding category model. Robust initial pose estimation is obtained using random sample consensus (RANSAC) algorithm and the least squares algorithm. Finally, the accurate pose estimation result is obtained using the point-to-plane iterative closet point (ICP) algorithm. The evaluation results in the CV-Lab 3D synthesis and UWA real-scene datasets show that the proposed algorithm significantly improves the interior point probability in the local feature matching stage, thereby improving the robustness and efficiency of pose estimation,

收稿日期: 2022-01-20; 修回日期: 2022-03-04; 录用日期: 2022-03-24; 网络首发日期: 2022-04-04

基金项目: 国家自然科学基金(62073312)、辽宁省“兴辽英才计划”项目(XLYC2002055)

通信作者: *liuyaohua19@mails.ucas.ac.cn

particularly in applications with multiple instances in the scene, compared with the direct matching model and local features of all scenes for multi-target pose estimation.

Key words image processing; point cloud; instance segmentation; pose estimation; local feature matching; random forests

1 引言

3D 目标的 6 自由度(6DoF)位姿估计是计算 3D 目标在场景中的位置与姿态,是计算机视觉中的一个重要问题,在工业机器人的自动上下料、服务机器人^[1]、自动驾驶环境感知^[2]以及增强现实(AR)等诸多领域有广泛的应用。然而实际应用场景中,由于传感器噪声、遮挡以及复杂背景等情况广泛存在,实现鲁棒性好效率高的多目标 6DoF 位姿估计仍然是一个挑战性问题。

目前,基于匹配点的 3D 目标位姿估计是广泛应用的方法,可以分为基于 2D-3D 对应点匹配的位姿估计和基于 3D-3D 对应点匹配的位姿估计。基于 2D-3D 对应点的位姿估计需要在离线阶段标注 3D 模型上 3D 关键点的位置,线上阶段在 2D 图像上进行 2D 关键点检测,与模型上的 3D 同名关键点匹配后,利用 PnP 算法^[3]根据相机成像模型得到场景中物体的 6DoF 位姿。常用的 2D 特征描述子有 SIFT^[4]、ORB^[5]、HOG^[6]等。这类方法对硬件要求低,且实时性好。但是,2D 关键点依赖于图像上的纹理特征,且容易受到物体所处环境光照等条件的影响^[7]。而且,工业零件等物体通常缺少纹理,因此这类方法的应用场景比较有限。3D-3D 对应点匹配法直接基于 3D 点云进行计算,离线阶段利用点云特征描述子对所有的模型点云提取局部特征。线上阶段利用相同的点云特征描述子提取场景的局部特征,利用特征匹配算法找到场景和模型上特征相似的对点,经过初始位姿估计算法得到初始位姿,最后通过迭代最近点(ICP)算法^[8]得到精确的位姿。其中常用的 3D 特征描述子有 SHOT^[9]、FPFH^[10]、Spin-Image^[11]等。

使用 3D-3D 对应点匹配的优势是对环境光线的鲁棒性更好,不依赖于物体表面颜色信息,而且点云局部特征对平移旋转具有不变性,因此更适于 3D 目标位姿估计任务。然而在点云局部特征匹配阶段,传统方法直接将模型上的局部特征与场景点的局部特征进行匹配,算法效率较低。与此同时,由于传感器数据噪声、场景中的模型遮挡、杂乱背景以及局部特征的类间相似性造成的不同类别之间错误匹配,会降低位姿估计鲁棒性和精度。

针对以上基于 3D-3D 对应点匹配的位姿估计中存在的问题,传统的方法是在已经产生大量类间错误匹配的情况下找到符合刚体一致性变换的匹配,通过随机采样一致(RANSAC)^[12]等算法剔除错误样本。在多目标位姿估计的场景下,为了获得鲁棒性通常需要大量的迭代,效率低下。

为了改进传统算法对于多目标位姿估计的不足,本文设计了点云实例分割算法,并在已知场景中实例目标的情况下进行多目标位姿估计,避免由于局部特征的类别区分度不足而造成类间错误匹配问题,而且能够缩小特征匹配的搜索空间,提高匹配阶段的鲁棒性。而且,相比直接使用模型遍历的策略进行多目标位姿估计,本文算法对于多目标场景的位姿估计无需遍历所有的模型,因此可以获得更高的效率。

2 算法原理

2.1 算法概述

传统的位姿估计方法直接匹配模型与场景的局部特征,然而由于不同类别的模型之间存在相似的局部几何结构,因此模型和场景的局部特征容易产生类间错误匹配的问题(如图 1 所示,中间的“龙”模型与周围场景产生了错误匹配),造成后续位姿估计算法的鲁棒性下降。为了解决这个问题,本文基于机器学习算法获得点云的实例分割,根据场景中不同点云实例的类别进行特征匹配,由此提高模型与场景点云局部特征匹配的内点概率,最终提高多目标 6DoF 位姿估计的效率和鲁棒性。

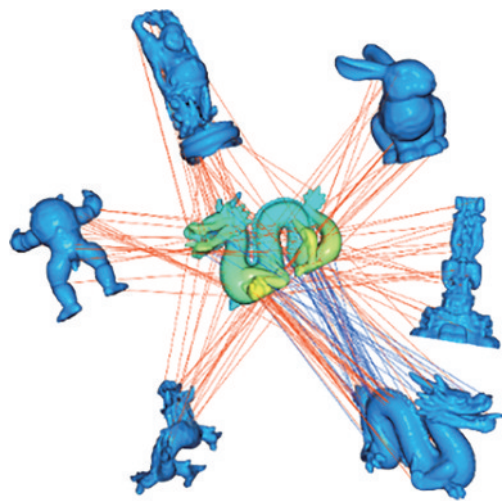
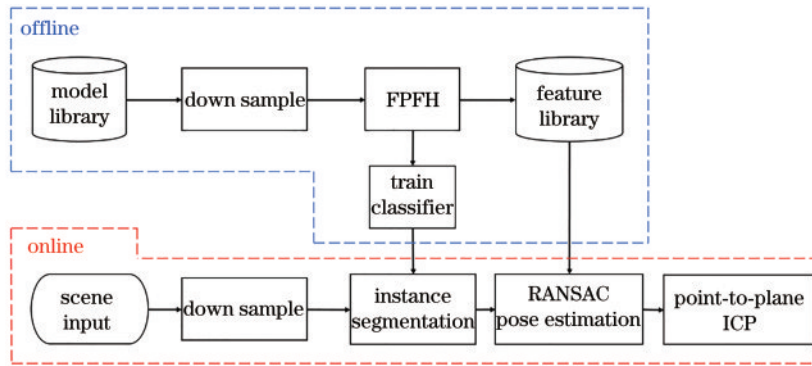


图 1 具有相似局部几何结构的多目标特征匹配

Fig. 1 Multi-target feature matching with similar local geometry

本文算法由线下阶段和线上阶段两部分组成:线下阶段利用模型库的点云局部特征训练机器学习分类器;线上阶段首先进行点云实例分割,然后基于 RANSAC 算法^[12]与最小二乘算法得到初始位姿估计,最后利用点到面(point-to-plane)ICP^[13]精配准,得到精确的位姿估计结果。算法框架如图 2 所示。



FPFH: fast point feature histogram; RANSAC: random sample consensus; ICP: iterative closet point

图2 提出算法框架

Fig. 2 Framework of proposed algorithm

2.2 点云实例分割

点云实例分割的目的是得到场景中所有感兴趣的目标实例,分为点云分割和点云簇分类两部分。为了降低计算复杂度,避免同一个位置被多次描述,首先对点云进行体素下采样^[14]。点云分割阶段,利用基于密度的聚类(DBSCAN)算法^[15]对采样后的场景进行基于密度的聚类得到多个点云簇。在分割后的点云簇上利用快速点特征直方图(FPFH)局部特征描述子进行局部几何特征描述。通过机器学习算法随机森林(RF)^[16]对点云簇的局部特征向量进行分类,得到所有点云簇的类别标签,从而完成点云的实例分割。

2.2.1 点云分割

DBSCAN是一种基于密度的聚类方法,在给定邻域半径之后,如果半径范围内包含大于设定个数的点,则所有邻接点视为同一类,对于点云数据,场景中散乱分布的3D模型可以通过DBSCAN得到不同的点云簇。

在利用深度相机(RGBD)或者激光雷达(LiDAR)等传感器得到的点云中,包含了较多噪声点。这些噪声往往孤立分布于场景中,即噪声点的密度与3D模型表面或者背景相比有较大差异。因此,利用DBSCAN能够有效检出并去除噪声点。

2.2.2 点云簇局部特征提取

为了获得分割后点云簇的类别标签,首先对其进行局部几何特征提取。尽管目前基于深度学习的点云分类和特征提取算法^[17]在特定的数据集上已经取得了较理想的效果,然而相比于传统的几何特征描述子,其模型训练和推理往往需要大量的训练数据和计算资源,很难应对仅有少量标注样本的应用。因此,本文选择FPFH点云局部特征描述子^[10]高效完成点云的局部特征提取。

点特征直方图(PFS)^[18]通过对关键点邻域内的所有点对进行互联,将邻域点的空间分布与法向量进行编码及量化,得到125维的特征向量,用于描述点云的局部几何特征。其计算复杂度为 $O(nk^2)$,其中 n 为所

有关键点的个数, k 为邻域点的个数。为提高计算效率,基于FPFH提出简化的点特征直方图(SPFH),FPFH首先在关键点 p 的 k 邻域计算SPFH,然后重新确定每个点的 k 邻域,并使用邻近点的SPFH加权计算得到关键点 p 的特征直方图。FPFH的计算原理为

$$F_{\text{FPFH}}(p) = F_{\text{SPFH}}(p) + \frac{1}{k} \sum_{i=1}^k \frac{1}{\omega_k} F_{\text{SPFH}}(p_k), \quad (1)$$

FPFH的计算复杂度为 $O(nk)$,特征向量为33维,可以实现实时计算,且易于实现快速匹配。

2.2.3 点云簇分类

FPFH点云局部描述子对近邻点的法向量以及空间坐标等几何信息进行描述,然而不同类别的模型上往往具有相似的几何结构,产生相似的局部特征向量。

通过对场景点云簇进行分类,即可从模型库匹配对应的模型。其基本的原理是将点云簇所有的局部特征聚合成点云簇的全局特征,此全局特征具有更大的感受野,从而具有更强的类别区分度。结合全局特征与机器学习分类器即可实现点云簇的分类。

本文基于RF算法实现局部特征向量分类。RF来源于决策树,决策树是机器学习中的常用分类器,为了应对复杂的数据,往往需要构造较深的决策树。然而,尽管更深的决策树能够以很低的偏差拟合复杂训练数据,但同时也会造成过拟合。RF通过平均多个决策树来降低预测的方差,从而提升分类器的泛化性能。机器学习算法通过数据驱动实现特征分类,避免了精细选择特定距离度量的工作。如图3所示。

点云局部特征分类包含RF分类器线下训练以及线上预测两部分。分类器训练阶段,对模型库的所有模型提取FPFH局部几何特征作为训练数据,以不同的类别作为训练分类器的标签。线上预测阶段,利用线下阶段训练的分类器对聚合后的点云簇局部特征进行分类。分类器给出置信度最高的类别即为点云簇的类别。

由于不同模型点云的数量差异较大,直接使用模型上所有的特征训练分类器会面临类别不平衡问题。

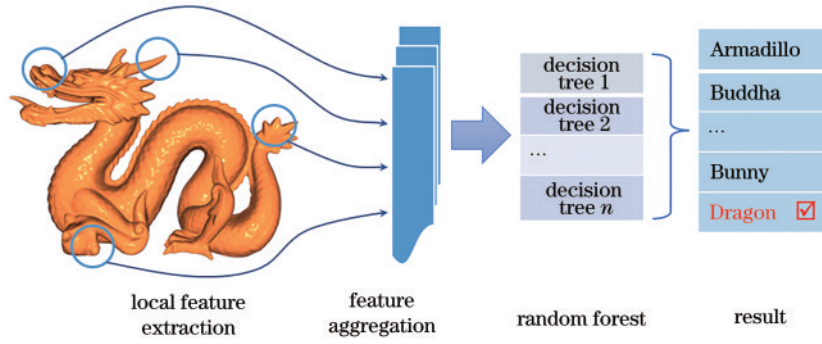


图 3 点云局部特征聚合与分类

Fig. 3 Point cloud local feature aggregation and classification

图 4(a)展示了使用 CV-Lab 3D 数据集下的所有模型直接训练分类器并预测模型类别的结果,其中 Dragon_vrip 的分类预测结果与 Statuette 的置信度非常相近(仅差 0.01),这在线上运行阶段容易造成鲁棒性的下降。因此,本文从每个模型全部的局部特征中

均匀采样出相同数量(统计每个模型的点云数量,以其中最少的点数作为采样数量)的局部特征,进行分类器训练。图 4(b)为特征均匀采样后进行分类器训练以及预测的结果,可见模型的预测置信度得到了明显的改善,解决了类别不平衡的问题。

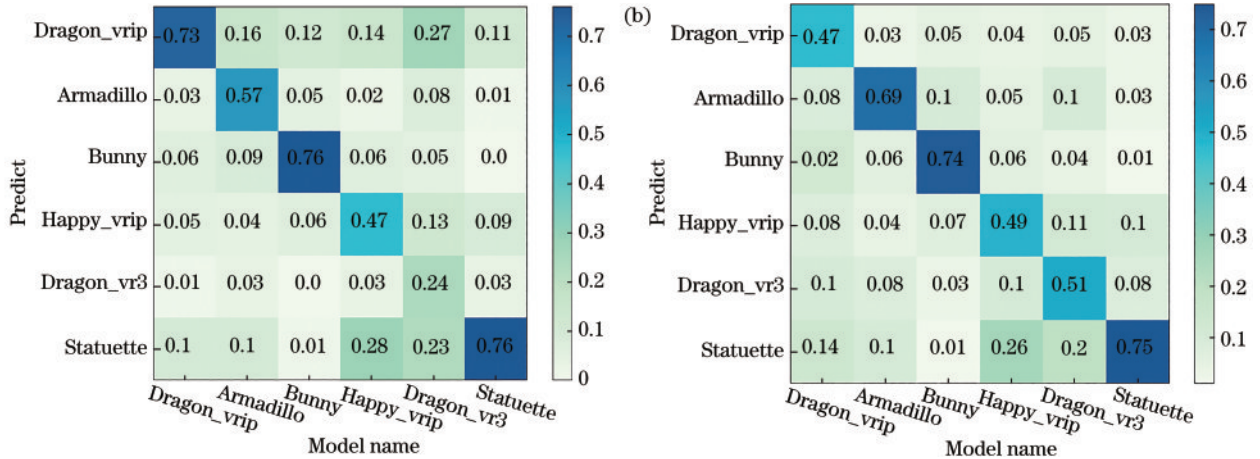


图 4 CV-Lab 3D 模型分类中的样本不平衡问题。(a)特征未经采样训练分类器;(b)特征采样后训练分类器

Fig. 4 CV-Lab 3D sample imbalance in model classification. (a) Training classifier without feature sampling; (b) training classifier after feature sampling

2.3 局部特征匹配与位姿估计

局部特征匹配的目的在于根据模型与场景的特征向量找到特征相似度最高的对应点集合。位姿估计则是利用匹配的点集求解 6DoF 位姿,包含初始位姿估计与精确位姿估计,其中初始位姿估计基于 RANSAC 与最小二乘估计得到满足刚体变换一致性的初始位姿,最后经过点到面 ICP^[13]得到精确的位姿估计结果。

2.3.1 点云局部特征匹配

由 2.2.2 节的 FPFH 局部特征提取,得到模型的特征集合 F_1 和场景的特征集合 F_2 ,其中每一个特征向量对应模型或场景中的一个 3D 点。本文使用近似近邻快速库 (FLANN) 算法^[19]进行特征匹配,得到模型特征 f 与场景中的最近邻特征 f'_1 和次近邻特征 f'_2 ,其中 $f \in F_1, f'_1 \in F_2, f'_2 \in F_2$ 。

为了提高匹配阶段的可靠性,将更有区分度的特征点作为正确匹配点,设定最近邻相似比 (NNSR)^[4]

进行匹配筛选,即最近邻特征与次近邻特征之间的关系应当满足:

$$1 - \frac{\|f - f'_1\|_{L_2}}{\|f - f'_2\|_{L_2}} \geq t_{NNSR}, \quad (2)$$

根据式(2)完成特征匹配筛选后,通过匹配的特征点索引可以得到模型与场景的 3D 匹配点集合 $\{P, Q\}$,用于位姿估计。

2.3.2 初始位姿估计

初始位姿估计的目的在于根据对应点集合 $\{P, Q\}$ 得到模型在场景中的初始位姿,保证 3.3 节精确位姿估计阶段的可用性(ICP 等算法需要一个很好的初始位姿)。为了保证类内特征匹配符合刚体变换一致性约束,对于场景中的每一个实例,本文使用 RANSAC 算法与最小二乘位姿估计算法得到初始位姿。

RANSAC^[12]是一种鲁棒的模型参数估计方法,思路是在对应点集合 $\{P, Q\}$ 上随机采样用于求解 6DoF

位姿的最小匹配点集 $\{p, q\}$, $p \in P, q \in Q$, 根据采样点集 $\{p, q\}$ 求解位姿变换, 并在全部的匹配点 $\{P, Q\}$ 上进行结果评估。经过多次迭代, 得到置信度足够高的位姿变换。记 $\{P, Q\}$ 中正确匹配点的数量为 T_{right} , 错误匹配点的数量为 T_{wrong} , 则其概率原理表示为

$$I_{\text{inlier}} = \frac{T_{\text{right}}}{T_{\text{right}} + T_{\text{wrong}}}, \quad (3)$$

$$1 - w = (1 - I_{\text{inlier}}^n)^k, \quad (4)$$

式中: w 为采样到理想匹配点对的概率; I_{inlier} 为正确的匹配点被采样到的概率 (范围是 0~1); n 为求解位姿所需要的最少点数。随着迭代次数 k 的增加, 得到理想匹配点的概率 w 趋近于 1, 由于理想采样点和理想位姿变换可以通过式(5)进行关联, 所以此时得到理想位姿变换 $[R, t]$ 的概率趋近于 1。

最小二乘位姿求解阶段, 设 $[R, t]$ 为模型变换到场景的旋转矩阵以及平移向量, 通过对 $\{P, Q\}$ 采样得到点集 $\{p, q\}$, 基于奇异值分解 (SVD) 求解^[20] 得到 $[R, t]$, 目标函数为

$$\arg \min_{R \in SO(3), t \in \mathbb{R}^3} \frac{1}{2} \sum_{i=1}^n \|(Rp_i + t) - q_i\|^2, p_i \in P, q_i \in Q, \quad (5)$$

令 $M = [R, t]$, 经过 RANSAC 算法的 r 次迭代, 得到解空间 $\{M_1, M_2, \dots, M_r\}$ 。

评估阶段, 利用 M 将模型上的匹配点变换到场景中, 对于每个模型上的匹配点 p_i , 求其到场景中匹配点 q_i 的距离, 即找到使目标函数 [式(6)] 最小的 $[R, t]$:

$$F(p, q) = \sum_{i=1}^n \|(Rp_i + t) - q_i\|^2. \quad (6)$$

2.3.3 精确位姿估计

ICP^[8] 算法通过查找两个点集中的最近点进行对应点匹配, 然后利用最小二乘法估计目标点云与场景

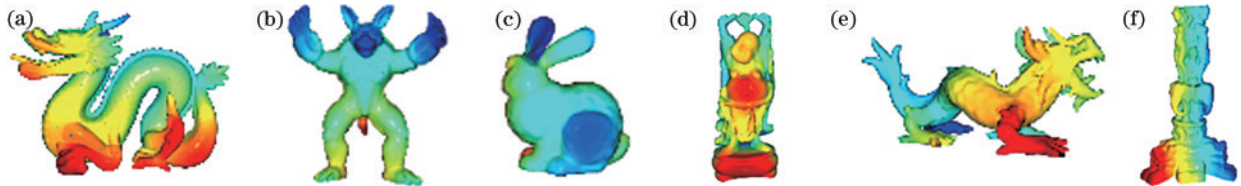


图 6 CV-Lab 3D dataset 1 中的点云模型。(a) Dragon_vrip; (b) Armadillo; (c) Bunny; (d) Happy_vrip; (e) Xyzrgb_dragon; (f) Statuette

Fig. 6 Point cloud models in CV-Lab 3D dataset 1. (a) Dragon_vrip; (b) Armadillo; (c) Bunny; (d) Happy_vrip; (e) Xyzrgb_dragon; (f) Statuette

UWA 数据集是利用 Minolta scanner 采集得到的单视角场景点云, 包含 5 个完整的模型以及 50 个场景, 并且提供了 4 个模型在每个场景中的真实位姿 (没有给出 Rhino 模型的真实位姿)。该数据集的场景中, 多个目标的堆放杂乱, 且模型之间存在大量遮挡 (每个场景中的模型, 平均仅有不足 25% 的部分可见^[23]), 对算法的鲁棒性要求较高。图 7 展示了该数据集所用模型。

点云之间的刚性位姿变换。点到面 ICP (如图 5 所示) 通过最小化场景切平面到模型点的距离作为优化目标。相比原始的 ICP 算法具有更快的收敛速度^[13]。

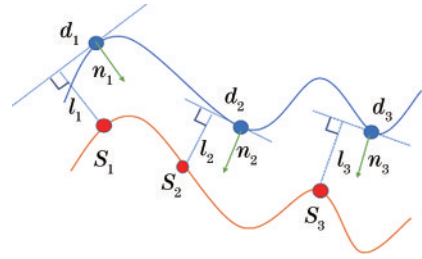


图 5 点到面 ICP 示意图

Fig. 5 Point-to-plane ICP diagram

令模型上的点 $s_i = (s_{ix}, s_{iy}, s_{iz}, 1)^T$, 场景切平面上的点 $d_i = (d_{ix}, d_{iy}, d_{iz}, 1)^T$, n_i 为切平面的单位法向量, 该算法的目标函数为

$$\arg \min_M \sum_i \|(M \cdot s_i - d_i) \cdot n_i\|^2, \quad (7)$$

当旋转量较小时, 旋转参数可以线性化近似, 经迭代求得精确的位姿变换矩阵 M 。

3 实验与结果讨论

3.1 数据集介绍

为了验证本文算法的有效性, 本文在 CV-Lab 3D^[9] dataset 1 和 UWA^[21] 公开数据集上进行多目标位姿估计算法评测。

其中, CV-Lab 3D dataset 1 数据集包含基于斯坦福 3D 扫描模型^[22] 制作的 45 个合成场景, 每个场景包含 3~5 个模型。合成场景中的每个模型都是完整的, 能够很好地测试多目标场景下点云局部特征匹配的性能。图 6 为该数据集使用的所有 3D 点云模型。

3.2 评测指标

3.2.1 点云局部特征匹配评测

点云局部特征匹配的精确率 (precision) 等价于 RANSAC 初始位姿估计阶段的内点概率, 由式(4)可知, 内点概率决定 RANSAC 所需的迭代次数, 从而决定初始位姿估计算法的运行效率和鲁棒性。

在基于 FLANN 算法计算得到的匹配点对中, 存在正确匹配点与错误匹配点, 其中正确的匹配点用 S_{TP}

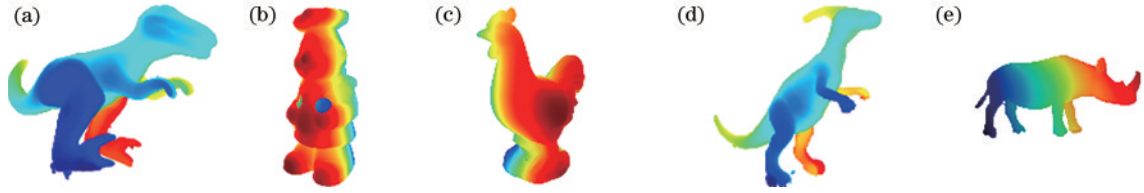


图 7 UWA 数据集中的点云模型。(a) T-rex; (b) Chef; (c) Chicken; (d) Parasaurolophus; (e) Rhino

Fig. 7 Point cloud models in UWA dataset. (a) T-rex; (b) Chef; (c) Chicken; (d) Parasaurolophus; (e) Rhino

表示,错误匹配点用 S_{FP} 表示,则点云局部特征匹配的精确率 $P_{precision}$ 为

$$P_{precision} = \frac{S_{TP}}{S_{TP} + S_{FP}} \quad (8)$$

实验中,匹配准确率的计算方法为:设置足够大的 RANSAC 迭代次数,以保证得到正确的初始位姿,利用该位姿变换对所有匹配点进行几何一致性检验(设定可以容忍的点对之间的欧氏距离作为阈值,对应点之间的欧氏距离小于阈值则视为符合几何一致性检验),符合几何一致性检验的点数视为 S_{TP} ,反之视为 S_{FP} 。

在 3.4.2 节,将分别给出在 CV-Lab dataset 1 数据集和 UWA 数据集上,每个模型在所有场景下的平均精确率。

3.2.2 位姿估计评测

本文以平均距离 (A_{dd})^[24] 定量评估位姿估计的结果。在给定了 3D 模型的真值位姿 $[R|t]$ 以及预测位姿 $[\tilde{R}|\tilde{t}]$ 之后, A_{dd} 对 3D 模型的点云进行变换,计算变换后模型上每个点对之间距离的平均值为

$$A_{dd} = \frac{1}{|N|} \sum_{x \in N} \|(Rx + T) - (\tilde{R}x + \tilde{t})\|, \quad (9)$$

式中: N 为 3D 模型点云上所有的点; $|N|$ 为模型点的数量。参考文献[24], 本文将 A_{dd} 与模型直径的 10% 进行比较,如果 A_{dd} 小于模型直径的 10%, 则视为 3D 目标位姿估计成功,记位姿估计成功的数量为 $S_{success}$, 位姿估计失败的数量为 S_{fail} 。以位姿估计的成功率作为位姿估计的评价指标,成功率 $R_{success}$ 的计算公式为

$$R_{success} = \frac{S_{success}}{S_{success} + S_{fail}} \quad (10)$$

3.3 实验参数设置

由于 CV-Lab 3D dataset 1 数据集与 UWA 数据集中的点云尺度差异较大,实验中,对不同的数据集分别设置参数。表 1 为 CV-Lab 3D dataset 1 合成数据集所用参数,表 2 为 UWA 真实场景数据集所用参数。

参数表中: Voxel size 为体素下采样算法的体素尺寸(单位: m); DBSCAN 算法的 Search radius 为密度聚类的搜索半径(单位: m), Min sample 为半径范围内包含的最少点数; RF 算法的 N estimators 为 RF 中决策树的数量, Max depth 为决策树的最大深度; FLANN 算法的 t_{NNSR} 为式(2)所用的阈值; RANSAC 算法的

表 1 CV-Lab 3D dataset 1 数据集所用参数

Table 1 Parameters used in CV-Lab 3D dataset 1

Algorithm	Parameter	Value
Down sample	Voxel size	0.006 m
	Search radius	0.008 m
DBSCAN	Min sample	6
	N estimators	100
RF	Max depth	10
	t_{NNSR}	0.85
FLANN	t_{NNSR}	0.85
RANSAC	Threshold	0.13

表 2 UWA 真实场景数据集所用参数

Table 2 Parameters used in UWA dataset

Algorithm	Parameter	Value
Down sample	Voxel size	0.40 m
	Search radius	0.55 m
DBSCAN	Min sample	5
	N estimators	100
RF	Max depth	10
	t_{NNSR}	0.85
FLANN	t_{NNSR}	0.85
RANSAC	Threshold	12

Threshold 为内点阈值。

3.4 实验结果

为了对比多类别场景下的性能提升, 本文利用多目标遍历的策略设计了基线算法, 也即, 对于每个场景, 遍历模型库中的所有模型点云, 分别将每一个模型的局部特征与场景中所有点云的局部特征进行匹配。利用 RANSAC 算法与最小二乘算法估计初始位姿, 如果内点概率大于设定的阈值, 就认为场景中存在该模型, 并且进入下一阶段的精配准, 得到模型最终的位姿。

3.4.1 改进效果可视化

为了直观对比改进前与改进后点云局部特征匹配的鲁棒性差异以及位姿估计的准确性, 本节从 CV-Lab 3D dataset 1 与 UWA 数据集中选出有代表性的场景进行可视化分析。其中, CV-Lab 3D dataset 1 选择场景 17, 该场景包含 4 个具有复杂局部几何结构的模型[如图 8(a)所示]。UWA 数据集选择场景 30, 该场景包含 4 个模型, 且平均可见表面不足 50% [如图 9(a)所示]。

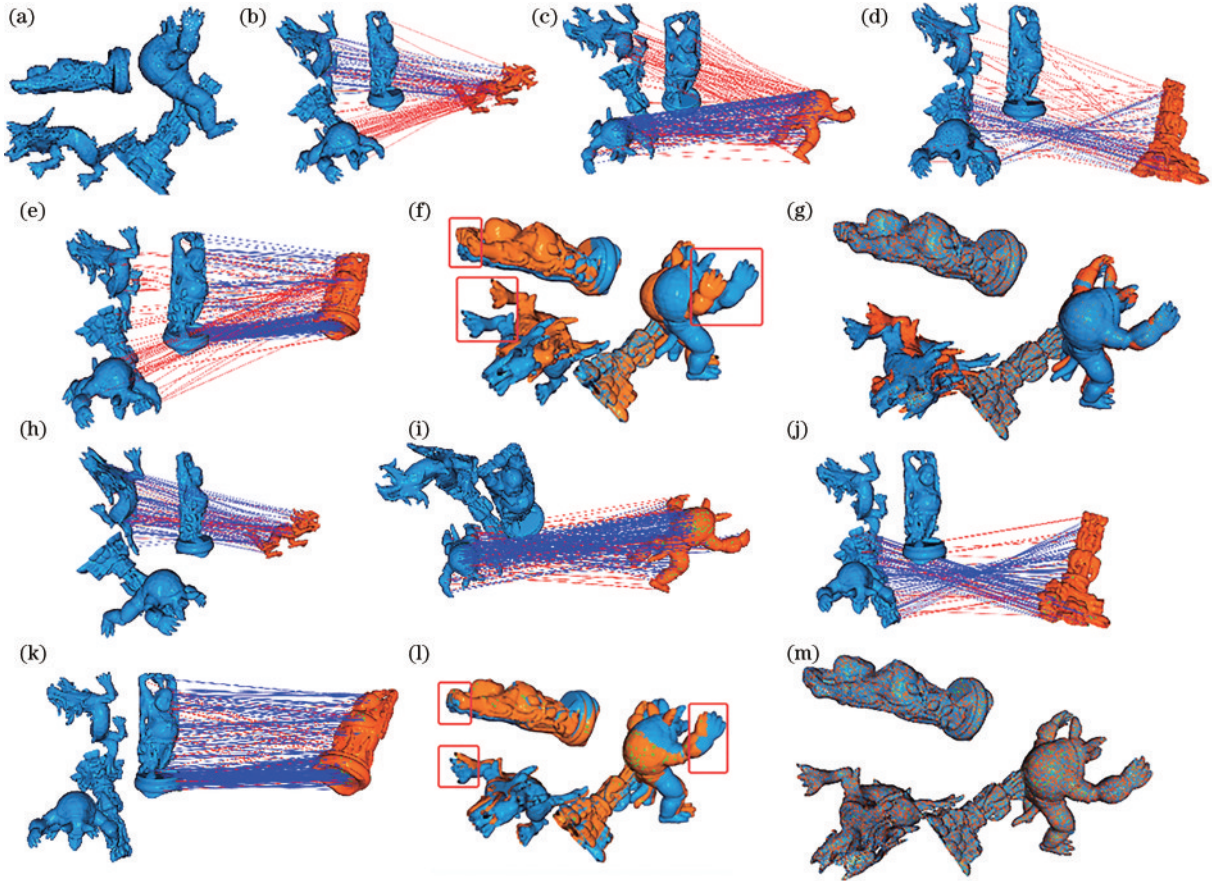


图 8 CV-Lab 3D dataset 1 场景 17。(a)场景中包含的模型;(b)~(e)改进前所有模型与场景的局部特征匹配;(f)改进前的初始位姿估计结果;(g)改进前的点到面 ICP 精确位姿估计结果;(h)~(k)提出算法改进后模型与场景的局部特征匹配;(l)提出算法改进后初始位姿估计结果;(m)提出算法改进后点到面 ICP 精确位姿估计结果

Fig. 8 CV-Lab 3D dataset 1 scene 17. (a) Models in scene; (b)~(e) local feature matching before optimization; (f) initial pose estimation result before optimization; (g) point-to-plane ICP pose refine result before optimization; (h)~(k) local feature matching after optimization using proposed algorithm; (l) initial pose estimation result after optimization using proposed algorithm; (m) point-to-plane ICP pose refinement result after optimization using proposed algorithm

图 8(b)~8(e)为基线算法(3.4 节所述)的点云局部特征匹配结果,图 8(h)~8(k)为本文改进后每个模型的局部特征匹配结果。从图中可以对比看出,未改进前的基线算法产生了大量的类间错误匹配,造成初始位姿估计阶段与真值的偏差较大[图 8(f)],即使经过点到面 ICP 也未能修正 Armadillo 和 Xyzrgb_dragon 模型的位姿[图 8(g)]。对比之下,由于本文实现了点云实例分割,场景中的每一个实例仅和对应类别的模型进行匹配,从而避免了产生类间错误匹配。在相同的 RANSAC 迭代次数(该场景设置为 100 次)下,改进后的算法实现了更加准确的初始位姿估计,从而保证点到面 ICP 得到准确的位姿[图 8(m)],其结果与真值位姿完全贴合。

图 9(b)~9(e)为基线算法在 UWA 数据集上的局部特征匹配结果。在 UWA 真实数据集上,除点云局部特征的类间相似性问题之外,另一个挑战性问题是场景中每个模型的可见部分仅有不足 50%。如图 9(b)的 T-rex 模型以及图 9(d)的 Chicken 模型均未能成功

匹配到场景中正确的类别上,导致位姿估计失败。经过本文算法改进之后,T-rex 以及 Chicken 能够匹配到正确的模型上,从而能够有效应对复杂的真实场景。

3.4.2 云局部特征匹配度量

根据 3.2.1 节所述点云局部特征匹配的度量方法,本节分别给出在 CV-Lab 3D dataset 1 和 UWA 数据集下,点云局部特征匹配的平均精确率。

表 3 为基线算法和本文改进后的算法在 CV-Lab 3D dataset 1 上,每个模型局部特征匹配的平均精确率统计结果。其中,基线算法对于 Dragon_vr3 模型的平均精确率仅为 0.444,本文改进后,该模型的平均精确率提升到了 0.816。该指标的提升能够有效提升 RANSAC 初始位姿估计阶段的鲁棒性。

表 4 为 UWA 数据集上的平均精确率统计结果。由于真实场景相比 CV-Lab 3D dataset 1 的合成场景更加复杂,基线算法部分模型的平均精确率不足 0.20。经过本文算法改进,所有模型的平均精确率均提升到 0.29 以上。

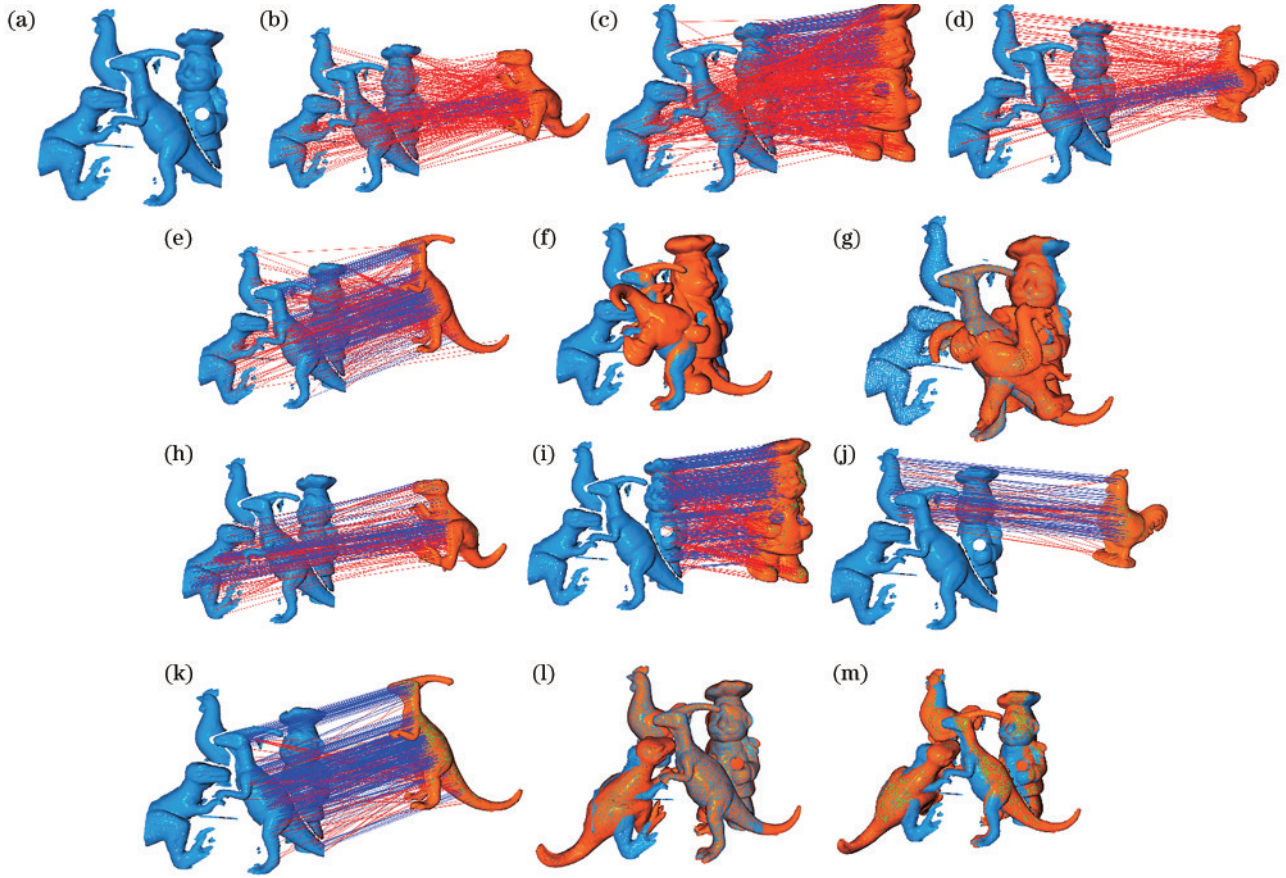


图9 UWA 场景 30。(a)场景中包含的模型;(b)~(e)改进前模型与场景间的局部特征匹配;(f)改进前初始位姿估计结果;(g)改进前点到面 ICP 精确位姿估计结果;(h)~(k)提出算法改进后模型与场景的局部特征匹配;(l)提出算法改进后初始位姿估计结果;(m)提出算法改进后点到面 ICP 精确位姿估计结果

Fig. 9 UWA scene 30. (a) Models in scene; (b)~(e) local feature matching before optimization; (f) initial pose estimation result before optimization; (g) point-to-plane ICP pose refine result before optimization; (h)~(k) local feature matching after optimization using proposed algorithm; (l) initial pose estimation result after optimization using proposed algorithm; (m) point-to-plane ICP pose refine result after optimization using proposed algorithm

表 3 CV-Lab 3D dataset 1 点云局部特征匹配的平均精确率
Table 3 Average accuracy of point cloud local feature matching on CV-Lab 3D dataset 1

Model	Baseline	Proposed
Armadillo	0.571	0.847
Bunny	0.563	0.690
Dragon_vrip	0.517	0.679
Happy_vrip	0.693	0.934
Dragon_vr3	0.444	0.816
Statuette	0.690	0.804

表 4 UWA 数据集点云局部特征匹配的平均精确率
Table 4 Average accuracy of point cloud local feature matching on UWA dataset

Model	Baseline	Proposed
T-rex	0.168	0.293
Chef	0.257	0.370
Chicken	0.236	0.556
Parasaurolophus	0.194	0.618

3.4.3 位姿估计度量

根据 3.2.2 的位姿估计评测指标,本节在 CV-Lab 3D dataset 1 上进行位姿估计定量评估。

由于基线算法以及改进后的算法使用了相同的精确位姿估计算法(点到面 ICP),因此,表 5 给出基线算法以及本文改进后算法的初始位姿估计结果。从定量评测结果可见,除 Statuette 模型的位姿估计准确率有 0.08 的下降以外,其余模型的位姿估计性能均得到了提升。其中,复杂模型 Dragon_vr3 的准确率提升了

表 5 UWA 数据集位姿估计准确率

Table 5 Pose estimation precision on UWA dataset

Model	Baseline	Proposed
Armadillo	0.909	1.000
Bunny	0.852	0.964
Dragon_vrip	0.844	0.853
Happy_vrip	0.750	0.893
Dragon_vr3	0.786	0.957
Statuette	0.880	0.808

0.171。该结果证明了本文提出算法在复杂场景下的有效性。

4 结 论

提出了一种点云实例分割算法,用于提高多目标 6DoF 位姿估计的鲁棒性。相比通过模型遍历等策略进行多目标位姿估计,所提算法首先对场景点云进行实例分割,从而避免点云局部特征的类型错误匹配问题,有效提高局部特征匹配的效率和鲁棒性。位姿估计阶段,采用 RANSAC 算法进一步解决由于点云局部特征的类型相似性导致的错误匹配问题,得到符合几何一致性的变换。最终通过点到面 ICP 得到高精度位姿估计结果。在 CV-Lab 3D dataset 1 的合成场景以及 UWA 实际采集场景上的实验结果表明,相比直接匹配原始的点云局部特征,所提算法在多实例位姿估计场景下鲁棒性更好,具有较高的实用价值。而且,在已经获得场景实例的情况下更容易实现多目标位姿估计的并行化,从而应用于机器人分拣、自动驾驶等需要高速 3D 感知的场景。

参 考 文 献

- [1] 党宏社, 陆馨蕊, 孙俊龙, 等. 基于视觉的取药机器人系统设计[J]. 激光与光电子学进展, 2021, 58(2): 0210009.
Dang H S, Lu X R, Sun J L, et al. Design of vision-based drug-grabbing robot system[J]. Laser & Optoelectronics Progress, 2021, 58(2): 0210009.
- [2] 胡杰, 刘汉, 徐文才, 等. 基于 3D 激光雷达的道路障碍物目标位姿检测算法[J]. 中国激光, 2021, 48(24): 2410001.
Hu J, Liu H, Xu W C, et al. Position detection algorithm of road obstacles based on 3D LiDAR[J]. Chinese Journal of Lasers, 2021, 48(24): 2410001.
- [3] 张慧娟, 熊芝, 劳达宝, 等. 基于 EPNP 算法的单目视觉测量系统研究[J]. 红外与激光工程, 2019, 48(5): 0517005.
Zhang H J, Xiong Z, Lao D B, et al. Monocular vision measurement system based on EPNP algorithm[J]. Infrared and Laser Engineering, 2019, 48(5): 0517005.
- [4] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [5] Rublee E, Rabaud V, Konolige K, et al. ORB: an efficient alternative to SIFT or SURF[C]//2011 International Conference on Computer Vision, November 6-13, 2011, Barcelona, Spain. New York: IEEE Press, 2011: 2564-2571.
- [6] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 20-25, 2005, San Diego, CA, USA. New York: IEEE Press, 2005: 886-893.
- [7] 张庆鹏, 曹宇. 室内场景下弱纹理物体 3D 重建算法的研究[J]. 激光与光电子学进展, 2021, 58(8): 0810017.
Zhang Q P, Cao Y. Research on three-dimensional reconstruction algorithm of weak textured objects in indoor scenes[J]. Laser & Optoelectronics Progress, 2021, 58(8): 0810017.
- [8] Besl P J, McKay N D. Method for registration of 3-D shapes[J]. Proceedings of SPIE, 1992, 1611: 586-606.
- [9] Tombari F, Salti S, di Stefano L. Unique signatures of histograms for local surface description[M]//Daniilidis K, Maragos P, Paragios N. Computer vision-ECCV 2010. Lecture notes in computer science. Heidelberg: Springer, 2010, 6313: 356-369.
- [10] Rusu R B, Blodow N, Beetz M. Fast point feature histograms (FPFH) for 3D registration[C]//2009 IEEE International Conference on Robotics and Automation, May 12-17, 2009, Kobe, Japan. New York: IEEE Press, 2009: 3212-3217.
- [11] Johnson A E. Spin-images: a representation for 3-D surface matching: CMU-RI-TR-97-47[R]. Pittsburgh: Carnegie Mellon University, 1997.
- [12] Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. Communications of the ACM, 1981, 24(6): 381-395.
- [13] Low K L. Linear least-squares optimization for point-to-plane ICP surface registration[R]. Chapel Hill: University of North Carolina, 2004, 4(10): 1-3. doi=10.1.1.116.7292&rep=rep1&type=pdf.
- [14] 张彬, 熊传兵. 基于体素下采样和关键点提取的点云自动配准[J]. 激光与光电子学进展, 2020, 57(4): 041008.
Zhang B, Xiong C B. Automatic point cloud registration based on voxel downsampling and key point extraction[J]. Laser & Optoelectronics Progress, 2020, 57(4): 041008.
- [15] Schubert E, Sander J, Ester M, et al. DBSCAN revisited, revisited: why and how You should (still) use DBSCAN[J]. ACM Transactions on Database Systems, 2017, 42(3): 19.
- [16] Freund Y, Schapire R E. A decision-theoretic generalization of on-line learning and an application to boosting[J]. Journal of Computer and System Sciences, 1997, 55(1): 119-139.
- [17] Charles R Q, Hao S, Mo K C, et al. PointNet: deep learning on point sets for 3D classification and segmentation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 77-85.
- [18] Rusu R B, Blodow N, Marton Z C, et al. Aligning point cloud views using persistent feature histograms[C]//2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, September 22-26, 2008, Nice, France. New York: IEEE Press, 2008: 3384-3391.
- [19] Muja M, Lowe D G. Scalable nearest neighbor algorithms for high dimensional data[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(11): 2227-2240.
- [20] Sorkine O. Least-squares rigid motion using svd[J].

- Technical Notes, 2009, 120(3): 52.
- [21] Mian A S, Bennamoun M, Owens R A. Three-dimensional model-based object recognition and segmentation in cluttered scenes[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(10): 1584-1601.
- [22] Curless B, Levoy M. A volumetric method for building complex models from range images[C]//SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, August 4-9, 1996, New Orleans, LA, USA. New York: ACM Press, 1996: 303-312.
- [23] Buch A G, Kiforenko L, Kraft D. Rotational subgroup voting and pose clustering for robust 3D object recognition[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 4137-4145.
- [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[M]//Lee K M, Matsushita Y, Rehg J M, et al. Computer vision-ACCV 2012. Lecture notes in computer science. Heidelberg: Springer, 2013, 7724: 548-562.