DOI:10.11918/j.issn.0367-6234.201804210

一种非结构环境下目标识别和3D位姿估计方法

任秉银, 魏 坤, 代 勇

(哈尔滨工业大学 机电工程学院, 哈尔滨 150001)

摘 要: 为提高非结构环境下目标识别准确率和位姿估计精度,提出一种利用 Kinect V2 RGB-D 传感器并基于目标的 CAD 模型进行不同类型目标自动识别和 3D 位姿估计的新方法.利用虚拟相机获取目标 CAD 模型的深度图像,并将目标的模型转化为点云图,采用体素栅格滤波减少场景点云中的点数;利用点对特征描述子(PPF)作为 CAD 模型的全局描述子,并将相似的 PPF 划分成一组放进一个 hash 表,用于识别和定位目标,所有目标的 hash 表组成了 3D 模型数据库;利用基于投票策略的方法对不同类型目标进行检测识别和 3D 位姿估计,并采用位姿聚类的方法和 ICP 配准进行位姿修正,再通过奇异值滤波剔除误匹配位姿,从而提高位姿估计精度. 在虚拟机器人实验平台仿真环境中分析了 3 种管接头的识别率和位姿估计误差,结果表明:3 种管接头平均识别率 96%,位置误差<4 mm,姿态误差<2°,能够满足机械臂抓取要求. 将提出的方法与两种主流位姿估计方法进行了对比实验,结果表明,提出的方法无论是识别率还是 $F1$ 分数都要优于其他两种方法.

关键词: 目标识别;3D 位姿估计;非结构环境;CAD 模型;位姿聚类;ICP 配准

中图分类号: TP391 **文献标志码**: A **文章编号**: 0367-6234(2019)01-0038-07

A novel method of target recognition and 3D pose estimation in unstructured environment

REN Bingyin, WEI Kun, DAI Yong

(School of Mechatronics Engineering, Harbin Institute of Technology, Harbin 150001, China)

Abstract: To improve target recognition accuracy and pose estimation precision in unstructured environment, a novel method for automatic recognition and 3D pose estimation of different kinds of targets is presented based on the object CAD model and Kinect V2 RGB-D sensor. The depth image of object CAD model is obtained by a virtual camera, and it is converted into one point cloud. A voxel grid filter is utilized to reduce the number of points for the point cloud of the scene, and the point pair feature (PPF) is used as the global descriptor of CAD model. The similar PPFs are classified as the same group and put into one hash table to recognize and locate the targets. The hash tables of all targets compose the 3D model database. A voting scheme is adopted for different kinds of objects recognition and 3D pose estimation. Pose cluster and ICP registration are employed to refine 3D pose. The accuracy of 3D pose estimation is improved by filtering the mismatching poses. The recognition rate and the pose estimation error of three kinds of pipe joints are analyzed in the environment of virtual robot experimentation platform (V-REP). The simulation results show that the average recognition rate of three kinds of pipe joints is 96%. The position error is less than 4 mm and the orientation error is no more than 2°, which meet the requirement of manipulator grasping. The experiments of comparison with other mainstream methods are performed. The experimental results show that the recognition rate and the $F1$ score of the proposed method are superior to those of two other methods.

Keywords: target recognition; 3D pose estimation; unstructured environment; CAD model; pose clustering; ICP registration

机械臂抓取目标物的操作已广泛应用于数控机床上下料和物流分拣等自动化生产线和流水自动化作业中[1],而上述机械臂作业环境属于典型的非结构环境,其特点为不使用或者尽量少使用工装夹具,即周围环境和目标信息在不采用外部传感器感知的前提下是不固定、不可知、且不可描述的,在这样的工作环境中,机械臂交互的场景必然是高度混杂无序的,随机抓取任务无疑面临更大的挑战[2]. 然而,目标检测与识别和 3D 位姿估计则是机械臂成功抓取和放置目标的重要前提. 近些年来,随着视觉传感器硬件成本的降低,基于视觉的环境感知则是机

收稿日期: 2018-04-28
作者简介: 任秉银(1966—),男,教授,博士生导师
通信作者: 代 勇,daiyong@hit.edu.cn

械臂实现随机抓取操作的关键技术.

在非结构环境下,不同种类以及同种的目标物之间的无序交错堆叠,相互遮挡,给视觉传感器感知检测、识别目标,进而对其进行位姿估计带来非常大的困难[3]. 目前,在广泛应用于工业领域中的各种视觉传感器中,RGB-D 相机具有明显优势,它可以被看做是一个 RGB 彩色单目相机和一个主动投射结构光的深度相机的组合传感器,并且以不低于 30 帧/s 的帧率为表面无纹理的零件实时提供较为准确的深度信息. 因此,研究基于 RGB-D 传感器的目标检测和 3D 位姿估计具有重要意义和广泛应用前景.

国内外学者提出众多检测识别和位姿估计方法,但因受限于非结构环境的困难和难点,仍存在各种不足. Hinterstoisser 等[4]基于模板匹配法,通过从 3D 渲染模型、嵌入量化颜色梯度和表面的法向量中提取匹配模板,从而得到鲁棒的目标检测结果. Rios-Cabrera 等[5]提出一种扩展方法,通过用一个聚类在所有模板中找到最佳选择,同时具有更快的检测速度. Cai 等[6]和 Hodaň 等[7]通过级联和混序编码策略优化了匹配方法,改善了 3D 位姿估计精度. David 等[8]和 Herbert 等[9]提出了基于 2D 局部特征描述子的方法进行目标识别,这些关键点描述子不随光照和轻微几何变换的改变而变化. Rusu 等[10]用视点特征直方图(VFH) 计算了视点方向与物体点云法向量方向的夹角,并用直方图统计了夹角的分布数量. 由于 VFH 对遮挡并不鲁棒以及不能得到完整的位姿估计,所以 Aldoma 等[11-12]提出聚合视点特征直方图(CVFH),克服了之前的缺陷,并将其和局部关键点描述子以及形状颜色特征相结合. Xie 等[13]也报道了一个相似的多模态方法. 针对平面物体识别和位姿估计,Lai 等[14]提出了一个树结构,但是位姿估计受到限制,因为它只能估计物体位姿中的一个旋转自由度. 尽管上述方法可以有效识别物体位姿,但是还要取决于已知背景平面的分割. 所以,对于复杂混乱的背景并不鲁棒,严重依赖于复杂的点云分割算法. 另外,一些局部不变特征描述子相继被提出,如基于沿着一个点表面法线分布[15]、曲面曲率[16]、spin 图像[17]、SHOT 描述子[18]等. 尽管这些特征不随着物体变换而改变,但它们却经常对噪声和点云精度误差以及特征描述子参数敏感. Wohlhart 等[19]采用卷积神经网络 CNN 生成目标描述子,有效捕捉到目标属性和位姿. Crivellaro 等[20]提出一种训练 CNN 的方法来检测目标物体,并以一些控制点的 2D 投影形式预测位姿,和其他方法相比,这种方法对遮挡鲁棒,可以估计无纹理物体的位姿. 尽管许多位姿估计方法使用RGB-D相机获得了 3D 数据,但是,同时识别、定位不同类型的目标以及拓扑结构而尺寸不同的目标时,仍存在诸多困难.

针对在非结构环境下现有位姿估计方法存在的不足,本文提出一种基于RGB-D相机的目标识别和 3D 位姿估计的新方法,可以快速鲁棒识别、定位非结构环境下不同类型的目标. 首先,利用目标 3D CAD 数据离线生成模型数据库,采用体素栅格滤波过滤点云中点的数量,从而减小计算时间;其次,利用基于投票策略的匹配方法对目标进行位姿估计,并利用 ICP 算法对初始位姿进行修正;最后,通过虚拟机器人实验仿真平台分析了本方法的位姿估计精度和识别率,并与其他主流方法进行了对比. 本文提出的目标识别和位姿估计方法简单可行,计算成本低,能够很好应对非结构环境的特殊性及其特点.

1　目标识别与位姿估计的总体框架

本文提出的目标识别与位姿估计方法的总体框架如图 1 所示.

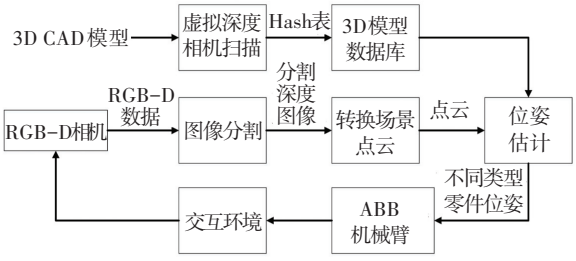

图 1　目标识别与位姿估计方法的总体框架

Fig.1　Framework of target recognition and pose estimation

离线阶段,使用虚拟相机和目标物的 3D CAD 几何模型数据生成目标的 3D 模型数据库;在线阶段,利用微软公司的 Kinect V2 深度相机拍摄目标的 RGB-D 图像并生成目标点云;最后通过基于投票策略的位姿估计系统识别不同种类目标的类型并估计它们的位姿.

2　目标物的 3D 模型数据库建立

使用 3D CAD 模型的优势在于当在线位姿估计时,可以快速创建 3D 模型数据库,建立 3D 模型数据库的流程如图 2 所示.

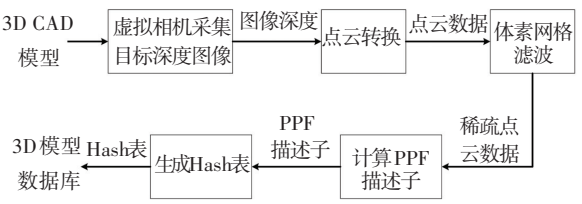

图 2　3D 模型数据库生成流程

Fig.2　Generation workflow of 3D model database

为了从 3D CAD 模型中获得有用数据来生成

3D 模型数据库,利用虚拟相机拍摄目标 3D CAD 模型的深度图像,进而将目标的深度图像转换成点云数据. 为了减少位姿估计计算时间,采用文献[21]提出的体素栅格滤波方法减少 3D CAD 模型点云中点的数量,然后估计这些点云数据的表面法线. 为了描述 3D CAD 模型,点云数据被转换成点对特征(PPF),点对特征用于将两个定向点配对,广泛用于基于形状的物体识别中[22]. 相似的点对特征被分组到同一个 hash 表中[23]. Hash 表是储存在 3D 模型数据库中的 3D CAD 模型的所有集合表示,3D 模型数据库则是用来估计 3D 位姿,从而最终实现识别和定位不同目标.

2.1 虚拟相机设置

为了从目标的 3D CAD 模型中生成深度图像,进而生成 3D 模型数据库,引入虚拟深度相机的概念,虚拟相机的建模仿真环境在图形库 Open GL 中进行,虚拟相机的设置见图 3,相机中心设置在虚拟球表面,并时刻指向球心.

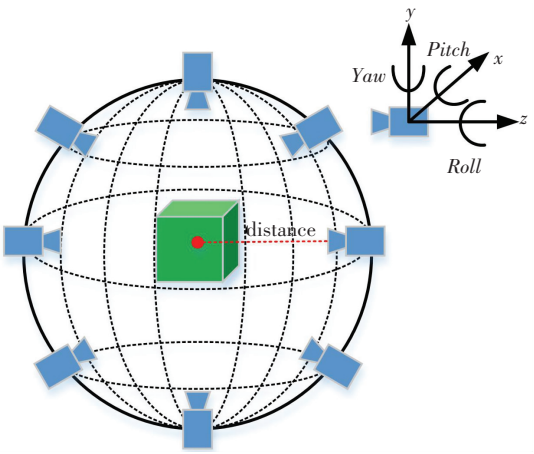

图 3 虚拟深度相机

Fig.3 Virtual depth camera

3D CAD 模型的中心位于虚拟球的中心,模型的深度图像是基于虚拟相机的位姿从虚拟球的不同位置和不同距离处拍摄的,设置虚拟相机的分辨率和 Kinect V2 分辨率相同,即彩色分辨率为 1 920×1 080,深度分辨率为 512×424. 虚拟相机在不同视角拍摄的深度图像融合生成 3D CAD 模型的点云. 虚拟深度图像的生成过程如图 4 所示,点云融合过程如图 5 所示.

虚拟相机拍摄的深度图像进而被转换成点云,对于给定的深度图像,相机位姿是已知的,深度图像的点云进而被转换到全局坐标系. 从相机坐标系到全局坐标系的点云变换可由以下公式计算得到:

$$P_g = \boldsymbol{T}_{g,k} P_k,$$

$$\boldsymbol{T}_{g,k} = [\boldsymbol{R}_c, \boldsymbol{t}_c].$$

式中:P_g 和 P_k 分别代表 CAD 模型在全局坐标系和相机坐标系下的点云,矩阵变换 $\boldsymbol{T}_{g,k}$ 把点云从相机坐标系变换到全局坐标系,$\boldsymbol{R}_c$ 和 $\boldsymbol{t}_c$ 分别代表虚拟相机在全局坐标系下的旋转矩阵和位置向量. CAD 模型在全局坐标系下的点云被储存后,虚拟相机依次移动到下一个点进行拍摄,并融合当前点云直到预先规划好的所有轨迹都完成. 通过点云融合,能够获取到更多的目标 3D CAD 模型的细节信息,从而生成目标物 3D 模型数据库.

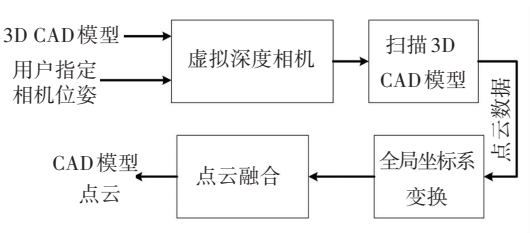

图 4 虚拟深度图像的生成流程

Fig.4 Generation workflow of virtual depth camera

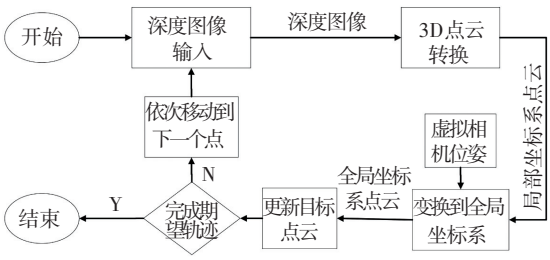

图 5 点云融合过程

Fig.5 Fusion process of the point cloud

2.2 体素栅格生成

由于生成 3D CAD 模型的点云时,获取了大量的点,包括离散点和稠密稀疏点,为了缩短计算时间,故采用体素栅格滤波减少点云中点的数量. 体素栅格滤波生成一个单位尺寸的 3D 体素网格,将对应的点云储存到每个体素栅格中,从而,原始点云被这些点云中的中心点所替换,如图 6 所示的过程. 原始点云被体素栅格滤波之后,虽然点的数量大规模减少,但点云形状却没有发生改变,从而保证了后期在线阶段更快更高效地进行位姿估计,却丝毫没有降低位姿估计的精度.

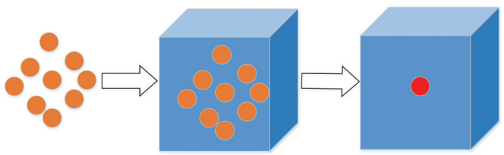

图 6 体素栅格处理

Fig.6 Voxel gird processing

2.3 Hash 表生成

Hash 表是一种数据结构,可以快速完成插入和

查找操作. 为了建立 3D CAD 模型的全局描述子，用点云来辨识点对特征描述子（PPF），并将相似的 PPF 划分成一组放进一个 hash 表中，如图 7 所示.

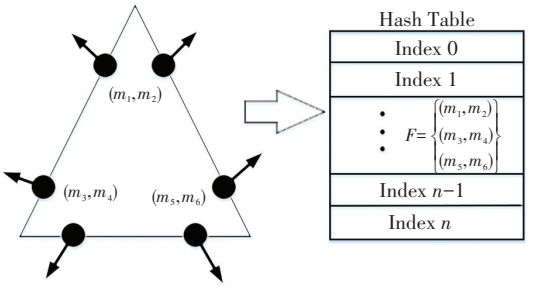

图 7　CAD 模型中具有相似的点对特征及其描述子储存在 hash 表相同位置

Fig.7　Point pairs having similar features and feature descriptors in the CAD model are stored in the same position in one hash table

Hash 表由 3D CAD 模型的点对特征组成，包含着目标的识别信息，PPF 描述子为

$$F(m_i,m_j)=(d,\varphi_1,\varphi_2,\varphi_3).$$

式中：m_i 和 m_j 是目标 3D CAD 模型上的点，d 是两个点之间的距离，φ_1 和 φ_2 分别是各自法线和这两点确定的向量之间的夹角，φ_3 则是这两条法线之间的夹角. 一个目标的 hash 表被保存在一个数据库中，图 7 表示点对 (m_1,m_2)，(m_3,m_4)，(m_5,m_6) 在单个目标上具有相似的特征，这些相似点对被收集保存在 hash 表中同一位置. 3D 模型数据库包括所有目标的 hash 表，可用来识别并估计目标的 3D 位姿，从而识别和定位目标. 根据场景的 (m_1,m_2) 和特征描述子 $F_s(s_r,s_i)$ 来匹配 CAD 模型的特征描述子 $F_m(m_r,m_i)$.

3　目标识别与 3D 位姿估计

位姿估计系统总体架构如图 8 所示.

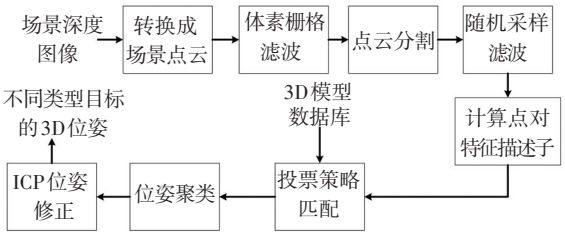

图 8　3D 位姿估计系统

Fig.8　3D pose estimation system

场景深度图像作为输入，不同目标的 3D 位姿作为输出. 提出的位姿估计方法分为离线阶段和在线阶段两部分：在离线阶段，生成 3D 模型数据库用来识别目标；在线阶段，位姿估计模块包括以下 7 个步骤：1）将场景深度图像转换成场景点云；2）使用体素栅格滤波减少场景点云数量；3）点云分割，依次获取各个目标的点云；4）估计场景点云表面法线；5）采用基于投票策略的匹配算法估计场景点云的 3D 位姿；6）采用位姿聚类算法移除不正确的具有高分的 3D 位姿；7）采用 ICP 算法修正 3D 位姿估计结果，从而提升位姿估计精度.

3.1　投票策略匹配

在线阶段，Kinect 传感器获取目标 RGB-D 图像并生成点云. 同样采用体素栅格滤波来减少场景点云的数量，然后估计场景点云表面法线 s_i，随机采样滤波从场景点云生成一系列参考点 s_r. 在参考点处逐一单独检测所有场景点云，特征描述子 $F_s(s_r,s_i)$ 和参考点 s_r 以及场景点云 s_i 构成一对. 为了滤除明显的点对特征描述子，把参考点 s_r 和场景点云 s_i 决定的法线夹角和距离定义合适的阈值，阈值选择依赖于目标尺寸和测试. 在本文中，距离和法线夹角的阈值分别设定为 12 mm 和 10°，这样的阈值设定可以有效地减少搜索区域和计算时间. 特征描述子 $F_s(s_r,s_i)$ 被用来搜索 hash 表中具有相似距离和法线方位的点对 (m_r,m_i). 点对 (m_r,m_i) 在局部坐标系中定位场景点云 (s_r,s_i) 如图 9 所示.

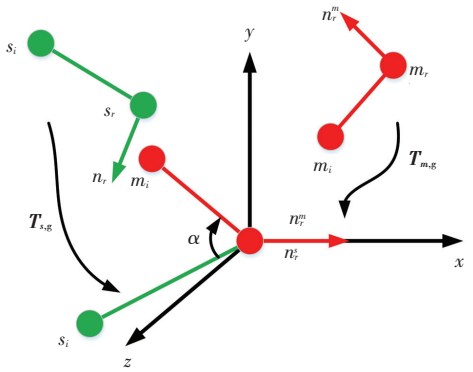

图 9　坐标系变换过程

Fig.9　Coordinate system transformation process

局部坐标系到场景坐标系的变换由下式变换得到[22]：

$$s_i=\boldsymbol{T}_{s,g}^{-1}\boldsymbol{R}_x(\alpha)\,\boldsymbol{T}_{m,g}m_i.$$

式中变换矩阵 $\boldsymbol{T}_{s,g}$ 和 $\boldsymbol{T}_{m,g}$ 把点 s_r 和 m_r 变换到局部坐标系的原点处，并且定位它们的表面法线 n_r^s 和 n_r^m 平行于 x 轴. 旋转矩阵 $\boldsymbol{R}_x(\alpha)$ 将模型方位点 m_i 绕着 x 轴转动 α 角度，从而定位场景方位点 s_i，每个 (m_r,α) 都是目标的识别和定位结果. 最后，建立一个由 r 和 α 定义的二维空间储存表，r 表示 CAD 模型点的数量，α 表示一系列离散角度，在局部坐标系 (m_r,α) 中采用投票策略. 当采用投票策略匹配完所有的点 s_i 时，储存表中所有具有最高投票数的最大值构成了位姿估计结果.

3.2 位姿聚类

投票策略匹配的结果是一系列 3D 位姿,同时包含关联着投票数的目标类型. 为了提高位姿估计结果的精度,本文采用文献[23]提出的位姿聚类程序来辨识投票策略结果中的聚类. 所有的 3D 位姿都聚合在一起以便在一个聚类中的所有位姿都彼此相似. 一个位姿聚类的分数是在一个聚类中所有位姿投票数的总和,当投票策略匹配上该位姿聚类时,这些投票数则会增加. 最后,辨识出具有最大投票数的位姿聚类,对聚类中的位姿求平均值,从而得到最终位姿结果. 位姿聚类的方法提高了采用基于投票策略匹配得到的 3D 位姿估计的精度.

3.3 ICP 位姿修正

点云配准是指将两个不同视点的点群三维数据整合到一个统一的坐标系的过程. 迭代最近点法(iterative closest points, ICP)是一种点集对点集的配准方法,其理论依据为:根据某种几何特性对数据进行匹配,并设这些匹配点为假想的对应点,然后根据这种对应关系求解运动参数,再利用这些运动参数对数据进行变换,并利用同一几何特征确定新的对应关系,重复上述过程进行迭代,使得数据中的重叠部分充分吻合.

采用位姿聚类获得一系列粗略的 3D 位姿,利用 ICP 配准算法来修正这些位姿,ICP 算法可以最小化两个点云之间的距离. 识别出的 CAD 模型的位姿给出了当前粗略的位姿,然后采用 ICP 算法将识别出的 CAD 模型点云和场景点云进行配准. 在 ICP 位姿修正后,通过计算识别到的 CAD 模型和场景点云之间的平均距离,从而求得配准误差. 如果该配准误差很小,则接受修正后的目标 3D 位姿.

3.4 奇异值滤波

误匹配结果包括识别(分类)误差和位姿估计(定位)误差,提出用奇异值滤波来解决这两种设计方法上的误差. 当不同类型目标非常相似时,投票策略的方法可能会带来分类误差,从而导致错误的识别目标类型. 当物料盒中的目标相互重叠遮挡彼此时,ICP 算法则可能会产生误定位误差,从而导致错误的位姿估计结果. 为了过滤掉误匹配位姿结果,奇异值滤波的输入是场景点云和一系列识别 CAD 模型的点云. 定义适应度分数用来移除错误匹配结果,适应度函数定义如下[24]:

$$E^{s}=\frac{100}{q}\sum_{i=0}^{q-1}\begin{cases}1, & \text{if } \min_{j=0}^{r-1}\| o_i - w_j \| > d_T,\\ 0, & \text{otherwise.}\end{cases}$$

式中: E^{s} 为计算出的目标欧几里得分数, o_i 和 q 为识别 CAD 模型的点云及其数量, w_j 和 r 为场景点云及其数量, d_T 为距离阈值,用来判断是否两点定位足够准确. 当识别 CAD 模型上的点 o_i 和场景中最近点 w_j 的距离大于 d_T 时,点 o_i 则是一个奇异值,本文中距离阈值 d_T 设定为 8 mm. 从以上适应度函数可以看出,精确的目标位姿估计结果应该是低分数的,如果当目标位姿误匹配时,CAD 模型点云则会具有高分数.

4 仿真与实验

下面以 3 种管接头的自动识别与位姿估计为例,验证方法的有效性. 通过 Kinect V2 RGB-D 传感器采集到的包含目标物的场景图像如图 10 所示.

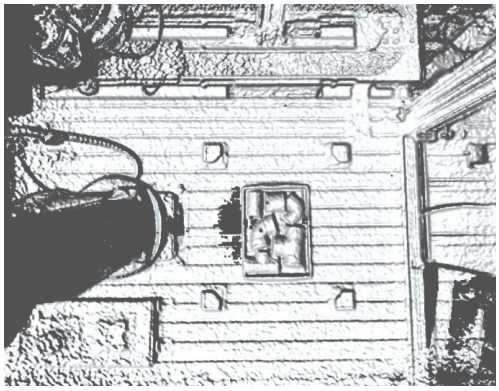

图 10 3 种管接头的场景分布

Fig.10 Scene distribution of three kinds of pipe joints

为了验证提出的 3D 位姿估计方法的有效性,在虚拟机器人实验平台(virtual robot experimentation platform,V-REP)中建立相同的虚拟场景,利用提出的方法对 3 种目标进行识别和位姿估计.

将利用 SolidWorks 软件建立的 3 种目标的 CAD 模型另存为 STL 格式,并直接导入到 V-REP 中;虚拟场景则在 V-REP 中建立,并且检测位姿估计的误差和识别率,虚拟深度相机用来采集虚拟场景中 3 种目标的深度图像,位姿估计方法在 V-REP 中以插件的形式编写程序,虚拟仿真环境如图 11 所示.

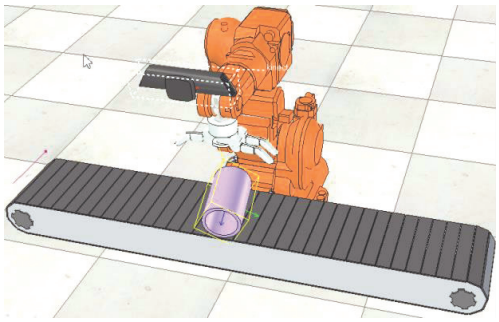

图 11 V-REP 虚拟仿真环境

Fig.11 Virtual simulation environment in V-REP

在 V-REP 仿真中，位姿估计程序用来验证根据场景点云计算出来的目标位姿的精度. 体素栅格尺寸的选择依赖于抓取目标的尺寸，本文选取的 3 种目标的尺寸见表 1.

表 1　3 种管接头的尺寸

Tab.1　Dimensions of three kinds of pipe joints　mm

目标类型	长度	宽度	高度
三通管	103.54	48.34	48.34
直角弯管	83.28	48.00	48.00
直通管	52.48	53.00	53.00

体素栅格的尺寸初步设定为 3 mm，体素栅格滤波将点云数量最小化，同时维持目标形状不变.

4.1　位姿估计误差分析

为了计算位姿估计的位置误差，识别目标的中心点从相应的 CAD 模型中心点提取；为了计算姿态误差，手动选择识别目标上的两个参考点，并计算其方向向量，选择相应的 CAD 模型上的相同点，计算这两个向量的角度获得姿态误差. 重复进行 50 次仿真，对随机放置的 3 种管接头目标，利用提出的方法估计位姿，位姿估计误差如表 2 所示，从表中可以看出位置误差<4 mm，姿态误差<2°.

表 2　位姿估计绝对误差

Tab.2　Absolute error of pose estimation

目标类型	X/ mm	Y/ mm	Z/ mm	旋转角度/(°)
三通管	1.23	1.82	1.56	1.23
直角弯管	2.52	3.78	3.83	1.54
直通管	0.88	1.25	1.58	1.82

4.2　目标识别率分析

3 种管接头目标随机放置，用虚拟相机获取其 100 张图像. 位姿估计系统识别 3 种目标的识别率见表 3 所示.

表 3　3 种目标识别率

Tab.3　Recognition rate of three kinds of targets

目标类型	总次数	失败次数	识别率/%
三通管	100	0	100
直角弯管	100	0	100
直通管	100	12	88
平均识别率			96

从表 3 中可以看出，三通管和直角弯管的识别率为 100%，直通管的识别率为 88%，3 种目标平均识别率 96%. 直通管的识别率相比三通管和直角弯管的识别率较低，主要是由于 Hash 表是由 3D CAD 模型的点对特征组成，包含着目标的识别信息，而直通管 CAD 模型的特征描述子 $F_m(m_r, m_i)$ 表达不明显.

4.3　与主流方法对比实验

为了进一步体现出本文提出方法的切实可行性和优越性，将提出的方法与文献[4]和[25]中的位姿估计方法进行了 102 次对比实验，采用识别率和 $F1$ 分数作为评价指标，$F1$ 分数是统计分析中一种衡量测试精度的指标，同时考虑了准确率(Precision)和召回率(Recall)，是准确率和召回率的加权平均，对比结果见表 4 和表 5.

$F1$ 分数计算公式如下：

$$F1 = \frac{2PR}{P + R}, \quad P = \frac{T_P}{T_P + F_P}, \quad R = \frac{T_P}{T_P + F_N}.$$

式中：P 表示准确率，R 表示召回率，T_P 表示目标被正确识别及位姿被正确估计，F_P 表示错误的识别结果或者不精确的位姿估计，F_N 表示漏检结果(即目标存在场景却未被识别出来). 由此可见，$F1$ 分数的范围是[0,1].

表 4　提出的方法和主流方法的识别率

Tab.4　Recognition rate of proposed method and mainstream methods

方法	三通管		直角弯管		直通管	
	成功次数	识别率/%	成功次数	识别率/%	成功次数	识别率/%
文献[4]方法	90	88.2	86	84.3	88	86.3
文献[25]方法	94	92.2	95	93.1	96	94.1
本文方法	98	96.1	96	94.1	89	87.3

表 5　提出的方法和主流方法的 $F1$ 分数

Tab.5　$F1$ score of proposed method and mainstream method

方法	三通管	直角弯管	直通管
文献[4]方法	0.512	0.487	0.687
文献[25]方法	0.623	0.752	0.513
本文方法	0.845	0.824	0.759

从表 4 和表 5 可以看出，本文提出的位姿估计方法无论识别率还是 $F1$ 分数都要优于其余两种方法.

5　结　论

1) 利用虚拟深度相机建立 3D CAD 模型的深度图像，然后从场景点云中利用点对特征描述子(PPF)作为全局描述子建立 3D 模型数据库，使用奇异值滤波剔除错误匹配位姿，采用基于投票策略的匹配方法以及 ICP 位姿修正实现了对物料盒中 3 种不同目标进行识别和 3D 位姿估计.

2) 在V-REP仿真环境中分析了位姿估计的位置和姿态误差以及识别率,位置误差<4 mm,姿态误差<2°,三通管和直角弯管的识别率为100%,直通管的识别率为88%,3种目标平均识别率96%. 真实对比实验验证了提出的目标识别与位姿估计方法的有效性,同时表明本文方法的识别率及$F1$分数都要优于两种主流位姿估计方法.

3) 本文提出的基于CAD模型的目标识别和3D位姿估计方法,利用Kinect V2 RGB-D传感器感知外界环境信息,为实现机械臂自主随机抓取不同目标和放置操作奠定了基础.

参考文献

[1] KAIPA K N, KANKANHALLI-NAGENDRA A S, KUMBLA N B, et al. Addressing perception uncertainty induced failure modes in robotic bin-picking[J]. Robotics & Computer Integrated Manufacturing,2016,42:17.DOI:10.1016/j.rcim.2016.05.002

[2] BRATANIC B, PERNUS F, LIKAR B, et al. Real-time pose estimation of rigid objects in heavily cluttered environments[J]. Computer Vision and Image Understanding, 2015, 141: 38. DOI: 10.1016/j.cviu. 2015. 09.002

[3] ASTANIN S, ANTONELLI D, CHIABERT P, et al. Reflective workpiece detection and localization for flexible robotic cells[J]. Robotics and Computer-Integrated Manufacturing, 2017, 44 (C): 190. DOI:10.1016/j. rcim.2016.09.001

[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]//Asian Conference on Computer Vision. Heidelberg:Springer,2012:548. DOI: 10.1007/978-3-642-37331-2_42

[5] RIOS-CABRERA R, TUYTELAARS T. Discriminatively trained templates for 3D object detection: a real time scalable approach[C]// IEEE International Conference on Computer Vision. Sydney:IEEE, 2014: 2048.DOI:10.1109/ICCV.2013.256

[6] CAI H, WERNER T, MATAS J. Fast detection of multiple texture-less 3-D objects[C]//International Conference on Computer Vision Systems. Heidelberg: Springer, 2013: 103. DOI: 10.1007/978-3-642-39402-7_11

[7] HODAŇ T, ZABULIS X, LOURAKIS M, et al. Detection and fine 3D pose estimation of texture-less objects in RGB-D images[C]// IEEE/RSJ International Conference on Intelligent Robots and Systems. Hamburg: IEEE, 2015: 4421. DOI: 10. 1109/IROS. 2015. -7354005

[8] DAVID G L. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2):91. DOI:10.1023/ B: VISI.0000029664.99615.94

[9] HERBERT B, ESS A, TUYTELAARS T, et al. Speeded-up robust features (SURF)[J]. Computer Vision & Image Understanding, 2008, 110(3): 346.DOI:10.1016/j. cviu.2007.09.014

[10] RUSU R B, BRADSKI G, THIBAUX R, et al. Fast 3D recognition and pose using the viewpoint feature histogram[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Taipei: IEEE, 2014:2155.DOI: 10.1109/IROS.2010.5651280

[11] ALDOMA A, VINCZE M, BLODOW N, et al. CAD-model recognition and 3DOF pose estimation using 3D cues[C]//IEEE International Conference on Computer Vision Workshops. Barcelona: IEEE, 2011:585.DOI: 10.1109/ICCVW.2011.6130296

[12] ALDOMA A, TOMBARI F, PRANKL J, et al. Multimodal cue integration through hypotheses verification for RGB-D object recognition and 3DOF pose estimation[C]//IEEE International Conference on Robotics and Automation. Karlsruhe: IEEE, 2013:2104. DOI: 10.1109/ICRA.2013.6630859

[13] XIE Z, SINGH A, UANG J, et al. Multimodal blending for high-accuracy instance recognition[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Tokyo: IEEE, 2013:2214. DOI: 10.1109/IROS.2013.6696666

[14] LAI K, BO L, REN X, et al. A scalable tree-based approach for joint object and pose recognition[C]//AAAI Conference on Artificial Intelligence. San Francisco: AAAI Press, 2011:1474

[15] STEIN F, MEDIONI G. Structural indexing: efficient 3-D object recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1992, 14(2): 125. DOI: 10.1109/34.121785

[16] CHITRA D, JAIN A K. COSMOS-A representation scheme for 3D free-form objects[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(10): 1115. DOI: 10.1109/34-625113

[17] JOHNSON A E, HEBERT M. Using spin images for efficient object recognition in cluttered 3D scenes[J]. IEEE Transactions on Pattern Analysis & amp, 2002,21(5):433.DOI:10.1109/34.765655

[18] TOMBARI F, SALTI S, STEFANO L D. Unique signatures of histograms for local surface description[J]. Lecture Notes in Computer Science, 2010, 6313:356

[19] WOHLHART P, LEPETIT V. Learning descriptors for object recognition and 3D pose estimation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston:IEEE, 2015:3109.DOI:10.1109/CVPR.2015.7298930

[20] CRIVELLARO A, RAD M, VERDIE Y, et al. A novel representation of parts for accurate 3D object detection and tracking in monocular images[C]// IEEE International Conference on Computer Vision. Santiago:IEEE,2015:4391.DOI: 10.1109/ICCV.2015. 499

[21] SKOTHEIM O, LIND M, YSTGAARD P, et al. A flexible 3D object localization system for industrial part handling[C]// IEEE/RSJ International Conference on Intelligent Robots and Systems. Vilamoura: IEEE, 2012: 3326. DOI: 10.1109/IROS.2012. 6385508

[22] DROST B, ULRICH M, NAVAB N, et al. Model globally, match locally: Efficient and robust 3D object recognition[C]// Computer Vision and Pattern Recognition. San Francisco: IEEE, 2010: 998. DOI: 10.1109/CVPR.2010.5540108

[23] CHOI C, TAGUCHI Y, TUZEL O, et al. Voting-based pose estimation for robotic assembly using a 3D sensor[C]// IEEE International Conference on Robotics and Automation. Saint Paul: IEEE, 2013:1724. DOI: 10.1109/ICRA.2012.6225371

[24] SQUIZZATO S. Robot bin picking: 3D pose retrieval based on Point Cloud Library[D]. [S.l.]: University of Padova, 2012

[25] BRACHMANN E, KRULL A, MICHEL F, et al. Learning 3D object pose estimation using 3D object coordinates[C]//European Conference on Computer Vision 2014. Amsterdam: Springer International Publishing,2014:536. DOI:10.1007/978-3-319-10605-2_35

(编辑 杨 波)