

DOI:10.11918/j.issn.0367-6234.201808133

3D 混杂场景中机械臂自主分拣小目标的方法

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:为解决机械臂在大小目标共存的3D混杂场景中无法利用3D视觉传感器直接感知分布于操作视场范围内的小目标这一难题,提出一种基于“固定安装的全局 Kinect 深度相机”与“安装在机械臂末端执行器上的移动相机(手眼相机)”相结合的视觉系统混合配置方法. 固定的全局 Kinect 深度相机用于感知并获取视场范围内的大目标点云,进而识别估计其位姿,然后借助路径规划技术引导机械臂到达大目标的上方,启动手眼相机近距离获取小目标的图像;离线阶段获取小目标的 CAD 模型,虚拟 2D 相机在以目标中心为球心的虚拟球表面的不同位姿和不同半径处拍摄目标的一系列二维视图,并且储存在目标的 3D 形状模板数据库中;在线阶段从真实手眼相机拍摄的场景图像中基于图像金字塔分层逐一搜索匹配,找到与目标模板相匹配的所有实例并计算其二维位姿,经过一系列转换后得到在相机坐标系下的初始三维位姿,应用非线性最小二乘法对其进行位姿修正. 由 ABB 机械臂和微软 Kinect V2 传感器以及维视图像公司的工业相机进行位姿估计精度实验和混杂目标分拣实验,利用棋盘标定板来测定目标真实的位姿. 实验结果表明,位置精度 0.48 mm,姿态精度 0.62°,平均识别时间 1.85 s,识别率达到 98%,远高于传统的基于特征和基于描述符的位姿估计方法,从而证明了提出方法的有效性和可行性.

关键词: 机械臂; 3D 感知; 小目标; 手眼相机; CAD 模型; 模板匹配; 自主分拣

中图分类号: TP391

文献标志码: A

文章编号: 0367-6234(2019)07-0042-07

A novel method for small object autonomous sorting of robotic manipulator in 3D clutter scene

REN Bingyin, WEI Kun, DAI Yong

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

Abstract: It is very difficult for a robotic manipulator to perceive and manipulate small objects directly using 3D visual sensors within its vision range in the scene where big and small targets are co-existed in 3D clutter scene. To solve the problem, a method for hybrid configuration of vision system based on fixed globally Kinect depth camera and fixed in robotic end effector moving camera (eye-in-hand camera) is proposed, in which the fixed globally Kinect depth camera is adopted to perceive and obtain the point clouds of big targets within its vision range, and their poses are recognized and estimated, which is utilized to guide the manipulator to move and arrive at big targets using path planning technology. An eye-in-hand camera is launched to capture the images of small object. In offline phase, the CAD model of a small object is created. A set of 2D view images are captured by a virtual 2D camera which is located at the surface of a sphere whose center is pointed into an object at different pose and radius, and stored in a database of 3D shape template of the object. In online phase, the scene image captured by a real eye-in-hand camera is explored and matched hierarchically one by one in details based on image pyramid to find all the instances matching with object templates and to compute their 2D poses. Initial 3D pose is obtained with respect to camera frame coordinate through a series of transformations. Rough pose is refined based on nonlinear least squares method. Experiments of pose estimation accuracy and industrial clutter objects sorting application are performed with ABB robotic manipulator, Microsoft Kinect V2 sensor and Micro Vision industrial camera. A checkerboard is employed to determine the true pose of the object. The results show that the position and orientation accuracy is 0.48 mm and 0.62°, respectively, and the recognition rate is 98% with average time 1.85 s, which is much higher than those of traditional feature-based and descriptor-based pose estimation methods.

Keywords: robotic manipulator; 3D perception; small object; eye-in-hand camera; CAD model; temple matching; autonomous sorting

近年来,3D 视觉传感器,尤其是低成本的消费级 RGB-D 传感器的应用大大提高了机械臂对周围

交互环境的 3D 感知能力,从而使得机械臂能够自主完成操作任务,如典型的机械臂对无序杂乱放置的目标物体实施抓取、搬运和装配作业^[1]. 在机械臂分拣任务中,经常会遇到形状和尺寸不同的目标混杂堆叠在一起. 对大小目标共存的混杂场景的感

收稿日期: 2018-08-27

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

知与识别是机械臂 3D 感知领域的一大难点, 绝大多数 RGB-D 传感器由于受限于自身深度分辨率的影响, 使得机械臂无法精确感知和识别其视场范围内的小目标物体, 从而导致机械臂无法从大小目标混杂物品中分拣出小目标物体。目前, 国内外学者们对上述场景中小目标的感知研究较少, 一般是研究机械臂对周围环境中大小相同或者相似的大目标的 3D 感知^[2-5]。识别和操作方法是借助 RGB-D 传感器获取场景点云, 然后基于离线创建的 CAD 模型的模板匹配与学习训练的方法, 在线对分割后的场景点云进行 ICP 配准, 从而估计出目标物体的三维位姿, 最终实现机械臂对目标的抓取、搬运和放置操作。为克服 RGB-D 传感器分辨率精度和视野范围的局限性, 一些学者们提出了结合手眼相机的混合配置方法, 借助单个相机拍摄的图像即可估计目标的位姿, 主要分为基于特征和基于描述符两种估计方法。基于特征的方法^[6-10]使用一些局部特征, 比如灰度值边缘, 直线的交集或者更复杂的组合提取基元特征, 这些特征对应于纹理边缘、几何边缘或者目标的角点。提取的特征用来匹配目标的几何特征, 直接从相应的点计算目标的三维位姿。然而, 背景的杂乱大大增加了提取目标局部特征数量的难度, 同时特征提取鲁棒性差, 对外界光照非常敏感, 尤其当目标部分被遮挡或者目标边缘不是清楚可见时。使这种基于局部特征的位姿估计方法并不适用于背景杂乱的工业零部件分拣场景。基于描述符的方法^[11-15]首先创建目标的人工视图, 然后确定其中特征点与从特征点的周围环境导出的差异变化描述符。基于这些描述符, 离线训练一个分类器。在线搜索阶段, 对从搜索图像导出的描述符进行分类, 建立模型与搜索图像之间的对应关系。基于描述符方法的最大优点是其运行时间与几何搜索的大小无关, 但仅局限于有纹理的目标识别, 并不适用于对弱纹理的工业零部件目标的识别和位姿估计。

针对 Kinect V2 传感器无法感知视场范围内的小目标以及无法进一步完成分拣操作的问题, 本文提出一种在 Kinect V2 传感器全局固定在场景的基础上, 结合移动手眼相机单目视觉的混合配置方法。当视场内的大目标处于深度传感器分辨率范围内时, 利用 Kinect 感知和识别目标, 机械臂进行大范围移动操作; 当视场内的小目标无法被 Kinect 识别时, 可以使用 Kinect 感知小目标周围的大目标, 进而能够引导机械臂从初始位姿靠近目标, 然后利用机械臂末端执行器的手眼相机去感知识别小目标, 通过离线阶段基于小目标三维 CAD 几何模型, 使用虚拟 2D 相机在不同位姿拍摄并生成其一系列基于

图像金字塔分层的二维模板, 并储存在目标的形状模型数据库中; 在线阶段从目标图像中搜索匹配与之形状对应模板的所有实例, 进而识别定位目标, 实现了仅仅使用单目相机拍摄的图像, 即可估计出目标在相机坐标系的三维位姿, 经过手眼标定转换到机械臂基坐标系的三维位姿后, 最终机械臂能够自主完成小目标抓取、移动和分拣等高精度操作。

本文提出的基于低成本的全局 RGB-D 传感器和局部手眼相机相结合的混合配置方法, 在工业现场环境中易于安装和部署, 对实现 3D 混杂场景中微小零部件的工业自动化柔性分拣、上下料和装配具有重要的意义。

1 问题描述

机械臂抓取分拣大小各异的目标混杂场景如图 1 所示, 目标物体包括三通管、直角弯管、直通管、水杯和象棋等物品。RGB-D 传感器采用微软公司的 Kinect V2 深度传感器, 将之固定在场景的右上角铝型材支架上, 拍摄场景目标的 RGB-D 图像并生成目标点云, 场景点云如图 2 所示。



图 1 机械臂抓取分拣混杂场景

Fig.1 Clutter scene for robotic manipulator grasping and sorting

从图 2 的场景点云图可以看到, Kinect 可以清楚地识别水杯和 3 种管接头这些大目标点云, 而象棋这些小物体由于受 Kinect 分辨率限制, 根本看不到它们的点云, 后续处理则不能被机械臂感知识别, 进而也就不能实现对这些小目标的自主抓取和分拣。

受眼在手上的相机可以近距离获得场景目标图像的启发, 故针对 RGB-D 传感器深度分辨率和视场范围的先天不足, 本文提出一种结合手眼相机混合配置的方法, 手眼相机被固定在 ABB 末端执行器上, 随机械臂移动, 近距离拍摄场景目标图像。

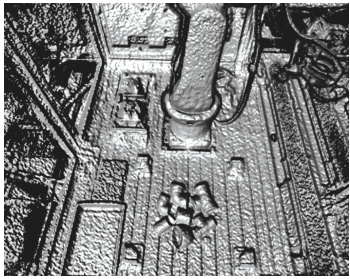


图 2 场景点云

Fig.2 Point cloud in scene

2 视觉系统混合配置方法总体框架

本文提出的视觉系统混合配置方法见图 3.

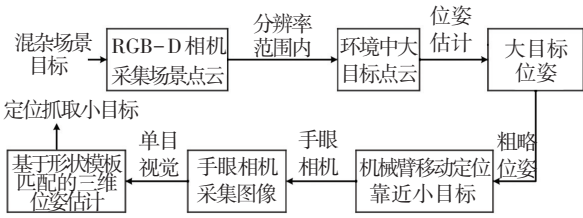


图 3 混合配置方法流程图

Fig.3 Flowchart of hybrid configuration method

当视场内的大目标(如水杯或者管接头)处于深度传感器分辨率范围内时,利用Kinect感知和识

别目标,机械臂进行大范围的移动操作;当视场内的小目标(如象棋等)无法被 Kinect 识别到时,使用 RGB-D 传感器感知并接近目标,然后利用机械臂末端执行器的手眼相机去感知识别小目标,提出基于目标三维 CAD 模型的三维模板匹配的方法,利用单目相机采集单张图像即可对目标进行三维位姿估计,从而完成目标的抓取、移动和放置任务. 其中,对 RGB-D 分辨率范围内的大目标的感知及其点云获取和位姿估计方法可以参阅文献[16].

3 基于形状的模板匹配的三维位姿估计方法

本文提出的基于形状的模板匹配的三维位姿估计方法由离线和在线两个阶段构成;离线阶段主要利用虚拟相机在不同半径的虚拟球表面的不同位姿处拍摄目标的一系列二维视图,并且储存在目标的 3D 形状数据库中;在线阶段从真实相机拍摄的场景图像中基于图像金字塔分层逐一搜索匹配,找到目标模板的所有实例及其二维位姿,经过一系列转换后得到在相机坐标系下的初始三维位姿. 最后,应用非线性最小二乘法对其进行位姿修正. 小目标识别与位姿估计方法总体流程如图 4 所示.

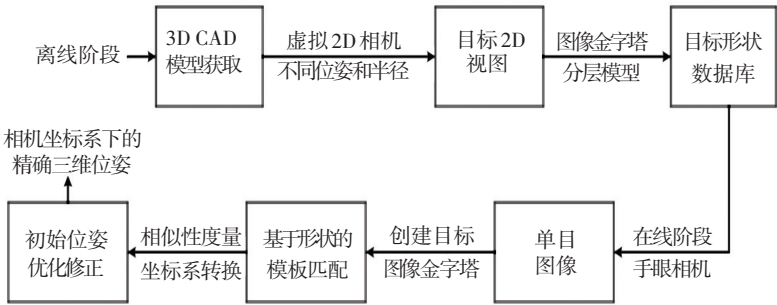


图 4 三维位姿估计方法流程图

Fig.4 Flowchart of three dimensional pose estimation method

3.1 离线 3D 形状模型生成

使用一个针孔模型的虚拟 2D 相机,相机内参数,包括焦距、畸变参数、像元尺寸以及主点等参数和真实所用相机参数完全相同. 假想一个虚拟球包围着目标 3D CAD 模型,目标模型处于球心,使用该虚拟相机从虚拟球的表面的不同位置拍摄目标图像,相机到目标模型的距离则由虚拟球的半球指定,取决于虚拟相机的位姿,根据球坐标系的定义,该位姿由球的 3 个坐标决定,即 (r, θ, γ) , 该过程示意图如图 5 所示. 根据虚拟相机投影自动创建了目标的一系列二维视图.

目标 2D 视图模型的创建需要使用文献[17]提出的一种相似性度量方法,对混杂场景中的遮挡、背景杂乱和非线性对比度的改变鲁棒性好.2D 模型由

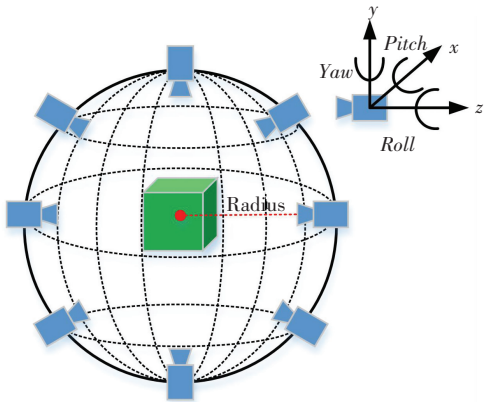


图 5 虚拟相机拍摄目标 CAD 模型的二维视图

Fig.5 Two dimensional view for the CAD model of an object captured by a virtual camera

相应梯度方向向量的多个边缘点组成,在此基础上,采用融合小边特征来增强背景变化的鲁棒性. 相似性度量是离线模型和在线搜索图像中相应的归一化边缘梯度方向点积的绝对值平均.

为了建立 2D 模型,目标的几何特征被用户指定位姿的虚拟相机投影到图像平面. 受文献[18]提出的图像彩色通道与模型表面几何特征的对应关系的启发,首先将目标 CAD 模型投影成三通道彩色图像,3 个通道分别代表目标表面法向量的 3 个分量,该彩色图像中测量的边缘幅值与目标的两个相邻面的法向量之间的三维空间中的角度直接相关. 假定两个相邻面的法向量分别为 $\mathbf{n}_1 = (x_1, y_1, z_1)^T$ 和 $\mathbf{n}_2 = (x_2, y_2, z_2)^T$, 当创建三通道图像时,假设第一个面用颜色 $(R_1, G_1, B_1) = (x_1, y_1, z_1)$ 涂画到图像上,而第二个面用颜色 $(R_2, G_2, B_2) = (x_2, y_2, z_2)$ 涂画到图像上,这两个投影面会在图像中产生垂直边缘. 因此,行方向的一阶导数为 $gr_R = gr_G = gr_B = 0$, 列方向的一阶导数 $gc_R = R_2 - R_1, gc_G = G_2 - G_1, gc_B = B_2 - B_1$. 彩色图像中的边缘幅值可以由颜色张量 $\mathbf{C}$ 的特征值计算获得:

$$\mathbf{C} = \begin{pmatrix} grr & grc \\ grc & gcc \end{pmatrix}.$$

对于三通道图像,则有

$$\begin{aligned} grr &= gr_R^2 + gr_G^2 + gr_B^2, \\ grc &= gr_R gc_R + gr_G gc_G + gr_B gc_B, \\ gcc &= gc_R^2 + gc_G^2 + gc_B^2. \end{aligned}$$

可以得到

$$\mathbf{C} = \begin{pmatrix} 0 & 0 \\ 0 & (R_2 - R_1)^2 + (G_2 - G_1)^2 + (B_2 - B_1)^2 \end{pmatrix},$$

则边缘幅值 A 是 $\mathbf{C}$ 的最大特征值的平方根,因此,

$$A = \sqrt{(R_2 - R_1)^2 + (G_2 - G_1)^2 + (B_2 - B_1)^2}.$$

在图像中计算的边缘幅值对应于两个法向量的差向量的长度. 两个单位法向量(长度为 1)跨越二维等腰三角形. 从边缘幅值可以很易导出两个法向量之间的角度 $\delta = 2 \arcsin(A/2)$.

将所获得的投影模型的彩色图像用作模型图像,通过颜色边缘提取进行扩展. 只有超过特定阈值的像素才会包含在模型中. 通常,模型的 3D 描述包含许多目标的真实图像中不可见的边. 这种边缘可由 CAD 软件的三角测量方法产生,该三角测量方法通过足够数量的平面来拟合近似弯曲的表面. 2D 模型是从图像金字塔级别上的三通道图像生成的. 如果在当前金字塔级别上创建的 2D 模型没有显示出足以区分模型与杂乱背景所必需的明显特征,则会自动淘汰. 3D 模型由若干金字塔级别上的多个

2D 模型组成,对于每个 2D 模型,存储相应的 3D 姿态. 最终,将所有 3D 模型的形状模板存储在一个数据库中,用于在线时进行搜索匹配.

3.2 在线模板匹配和位姿计算

在线阶段用生成的 3D 形状模型数据库在单目图像中匹配识别目标,从而计算出目标相对于相机坐标系下的三维位姿. 首先,从输入图像中创建目标的图像金字塔,匹配识别开始于最顶层的图像金字塔,在该层至少有一个合理的 2D 模型可以利用. 通过在离线视图中创建的 2D 模型和该层的图像金字塔计算一个相似性度量^[17],从而搜索到该层金字塔上所有 2D 模型. 该度量对遮挡、杂乱、对比度改变和局部极性改变鲁棒.

2D 模型在必要范围内旋转和缩放,并且在图像中缩放和旋转的 2D 模型的每个位置处计算相似性度量. 超过特定相似性阈值匹配的 2D 位姿(位置、旋转和缩放)被存储在匹配候选列表中. 在下一个较低层的图像金字塔中,所有树中没有父节点的 2D 模型的搜索方式与最顶层金字塔的视图相同. 此外,在先前金字塔上找到的匹配候选实例通过选择树中的所有子视图,并计算子视图的 2D 模型与当前图像金字塔之间的相似性度量来进行修正. 位置、旋转和缩放的范围可以限制为父匹配的近邻. 重复该过程,直到所有匹配候选者被查找到最底层的金字塔. 由于 2D 模型是在离线训练期间通过假设相机指向目标中心创建的,所以 2D 模型和成像目标通过二维的单应性矩阵相关联.

假设 x 是指向模型中心的相机投影 3D 模型到图像平面生成 2D 模型上的点, $\mathbf{K}$ 是相机标定参数矩阵, $\mathbf{R}$ 是相机旋转矩阵. 则旋转相机图像上的点 x' 可以表示为: $x' = \mathbf{K}\mathbf{R}\mathbf{K}^{-1}x = \mathbf{H}x$, 其中 $\mathbf{H}$ 是单应性矩阵.

未知旋转矩阵可以从图像投影模型的位置 $p = (r, c, 1)$ 齐次坐标计算得到,通过 $\mathbf{P} = \mathbf{K}^{-1}p$ 可以变换到三维空间. $\mathbf{R}$ 绕相机的 x 轴和 y 轴的旋转角度通过下式计算:

$$\begin{aligned} \alpha &= \arctan(P_y / \sqrt{P_z^2 + P_x^2}), \\ \beta &= \arctan(P_x / P_z). \end{aligned}$$

最终,使用子视图投影校正过的 2D 模型进行匹配. 在最顶层图像金字塔,由于没有先验知识,所以要彻底地穷举搜索,即在图像所有位置进行匹配. 作为匹配的结果,超过特定相似性度量的图像中的 2D 匹配则会得到 2D 位姿,即位置、旋转和缩放. 对于每次匹配,根据 2D 匹配位姿和关联的模型视图的三维位姿来计算与之对应的 3D 目标位姿. 模型视图的三维位姿可以表示为一个 4×4 齐次矩

阵 H_v , 它将目标坐标系下的点变换到相机坐标系. 2D 匹配的位姿则由位置 $p = (r, c, 1)$ 、旋转 γ 和缩放 s 给定. 必须修正矩阵 H_v 以便能正确表示 2D 匹配位姿.

首先应用 2D 缩放, 等同于目标和相机之间距离的相反各向同性缩放 S . 然后, 应用 2D 旋转, 等同于相机绕自身 z 轴的 3D 旋转 R_z . 最后, 图像中的位置等同于相机绕自身 x 轴和 y 轴的 3D 旋转. 两个旋转角度可以通过将位置变换为三维空间中的方向, 并随后以与上述类似的方式变换为旋转角度 α 和 β 来计算. 最终的可以描述目标相对于相机坐标系下的三维位姿的齐次变换矩阵为

$$H_{v,s,\gamma,p} = R_y(\beta) R_x(\alpha) R_z(-\gamma) S(1/s) H_v.$$

3.3 位姿修正

匹配计算得到的三维位姿的精度很大程度上受限于离线过程模型视图的采样和在线 2D 匹配过程中 2D 位姿的采样, 不利于背景混杂的工业应用检测识别.

本文应用非线性最小二乘法对初始匹配得到位姿进行修正. 利用位姿 $H_{v,s,\gamma,p}$, 3D 目标被投影到搜索图像中. 首先, 使用开源计算机图形库 Open GL 从分析投影的 3D CAD 模型边缘中移除隐藏线; 然后, 抑制表示两个相邻目标面之间的角度低于指定的最小面角的物体边缘线. 对每个视图离线计算以上述两个步骤导出的 3D CAD 模型边缘的可见性信息, 并将其存储在图像金字塔的分层模型中. 对可见的投影 CAD 模型的边缘采样为一个像素距离的离散点. 对于每个采样边缘点, 初始化局部搜索以便在采样边缘点的邻域中找到对应的亚像素精度的图像边缘点. 搜索仅限于垂直于投影 CAD 模型边缘的方向. 另外, 对于每个找到的可能对应关系, 计算垂直于投影的 CAD 模型边缘与图像梯度之间的角度差. 将低于阈值的角度差的对应关系被接受为有效的对应关系. 根据所有对应关系的角度差的中值, 鲁棒地计算该阈值.

非线性最小二乘法是以误差的平方和最小为准则来估计非线性静态模型参数的一种参数估计方法, 通常有搜索和迭代两类优化算法. 本文使用鲁棒的基于牛顿高斯迭代法与梯度最速下降法的非线性迭代优化 Levenberg-Marquardt (LM) 算法来获得修正后的三维位姿, 该算法对初值选取的依赖度较小, 能有效避免代价函数陷入局部极小值. MATLAB 已提供对应的算法优化拟合工具箱, 即非线性函数 lsqnonlin(fun, x0, lb, ub, options) 中把 options 选项设置为 'Levenberg-Marquardt' 算法即可. 优化过程中, 应用 6 个位姿参数直接最小化图像边缘点

到其对应的投影 CAD 模型边缘距离的平方, 优化后, 即可获得修正后的位姿参数. 随着迭代次数的增加, 三维位姿修正精度也在不断提高. 由于新的对应关系来自修正的位姿参数, 所以优化算法被集成在外部迭代循环中. 再次重新投影 CAD 模型, 并在每次迭代后重新计算对应关系. 假如两次迭代之间的对应关系(位姿参数)不再改变, 则终止迭代修正.

4 实验与结果分析

为了验证本文提出的 Kinect V2 深度传感器和手眼相机混合配置的基于形状匹配的三维位姿估计方法的有效性和可行性, 首先对小目标进行手眼相机位姿估计精度实验, 然后通过手眼标定将相机坐标系的位姿转换到机械臂基坐标系, 进而对混杂场景中的大小目标进行分拣实验.

4.1 实验设置与任务描述

实验场景及硬件组成如图 1 所示, 机械臂采用 ABB 公司的 IRB120, 负载为 3 kg, 重复定位精度为 ± 0.01 mm, 固定在全局环境的深度相机采用微软 Kinect V2 RGB-D 传感器, 其与机械臂相对位置关系已标定. 固定在机械臂末端执行器的手眼相机采用陕西维视图像公司, 型号 MV-VEM500SM, 分辨率为 500 万的工业相机, 像元尺寸 $2.2 \mu\text{m}$, 帧率 15 fps. 机械臂的运动带动手眼相机移动, 确保相机可以在任意位姿下近距离拍摄场景图像.

分拣任务: 机械臂根据大小目标的位姿估计结果, 抓取 3 种管接头、象棋和水杯, 最终将它们分别放入蓝色物料盒、象棋托盘和旁边的固定位置. 其中, 3 种管接头和水杯属于大目标, 用固定在右上角的 Kinect 传感器感知识别及位姿估计^[16]. 而象棋作为场景中的小目标, 则基于机械臂末端的手眼相机和本文提出的方法进行识别和三维位姿估计.

4.2 位姿估计精度实验

为了测试位姿估计的精度, 通过操纵示教器移动机械臂带动末端的手眼相机到达不同位姿, 分别拍摄 50 张场景目标图像. 把一个已知规格的标准的棋盘标定板(黑色标记点的中心在世界坐标系下的坐标已知)放到目标相邻处, 如图 6 所示. 根据标准的位姿估计算法, 可以计算出每一幅图像中标定板相对于相机坐标系的位姿, 通过计算目标相对于标定板的相对位姿. 最终, 可以计算出每一幅图像中目标的真实位姿.

采用本文提出的方法对每一幅图像中的目标进行位姿估计, 然后与上述真实位姿进行比较, 计算其相对标准差, 从而得出提出的位姿估计方法的精度. 为了证明本文的优越性, 同时和基于特征^[6]和基于

描述符^[11]的位姿估计识别方法的位置精度和姿态精度以及识别时间进行了对比,结果见表 1。由于小目标象棋具有丰富的纹理信息和特征,故非常适合采用基于特征和描述符的方法进行位姿估计。经过离线训练建立象棋的 3D CAD 模型,在线进行匹配时找到图像中所有象棋的实例,已经用红色标记框标记,从图 6 中可以看出,即使象棋被部分遮挡,同样可以识别出来。

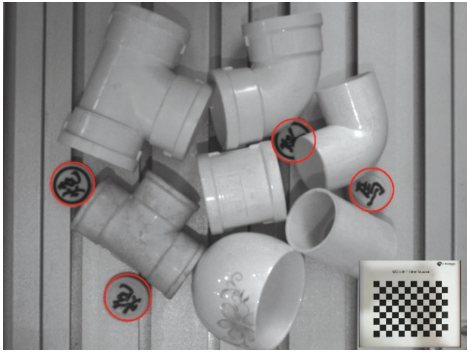


图 6 标定板测试目标的真实位姿

Fig.6 Calibration plate test of the real pose of the object
位姿估计精度测试实验环境为 Windows10+ Visual Studio 2010+Open CV 3.4.1。从表 1 的结果可以看出,本文提出的方法位置精度 0.48 mm,姿态精度 0.62°,识别率达到 98%,都远高于其余两种位姿估计方法,尽管识别时间相对较长,但都控制在 2 s 内,不影响该方法在实际工业中应用。

表 1 不同位姿估计方法的精度和识别率对比

Tab.1 Accuracy and recognition rate comparisons of different approaches to pose estimation

估计方法	位置精度/ mm	姿态精度/ (°)	识别时间/ s	成功 次数	识别率/ %
本文方法	0.48	0.62	1.85	49	98
基于特征方法 ^[6]	2.25	1.45	0.46	38	76
基于描述符方法 ^[11]	1.86	1.22	0.74	44	88

4.3 混杂大小目标分拣实验

Kinect 传感器首先感知混杂场景中的大目标,即 3 种管接头和水杯,对其进行位姿估计,进而依次抓取它们。大小目标的识别抓取可以交替进行,次序可以不固定,根据混杂场景(相互遮挡)的具体情况而定,即在完成一次大目标分拣后,记住当前位姿,借助路径规划技术^[19]机械臂可以从初始位姿自主到达小目标的周围,在保证手眼相机的工作物距前提下,启动手眼相机开始局部拍摄场景图像。

上述过程需要反复交替进行,因为场景中的大小目标是散乱堆叠在一起的,抓取一个目标后,其他目标可能会移动,从而导致目标位姿发生变化。利用本文提出的位姿估计方法,经过手眼标定转化后,

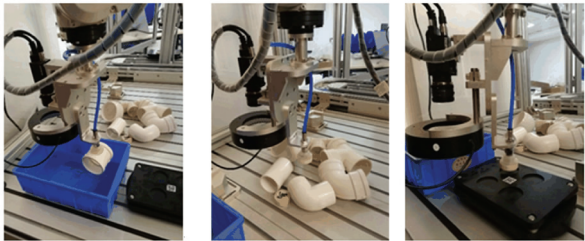
可以得到小目标(象棋)在机械臂基坐标系下的三维位姿,如图 7 (e)中,象棋的位姿为:

[432.5,60.8, - 40.2, 182.5°, 40.6°,172.4°] (mm)。

进而机械臂对象棋进行抓取、移动和放置,直至完成所有目标的分拣,整个分拣过程如图 7 所示,分拣结束场景如图 8 所示。通过在整个分拣工作节拍中设定合适的机械臂末端移动速度,混杂目标分拣总体耗时不超过 3 min,其中,4 个小目标分拣耗时 78 s,均控制在实际工业应用中可接受的时间范围内。



(a)抓取大目标(三通管) (b)抓取大目标(水杯) (c)抓取大目标(直通管)



(d)将直通管放入物料盒 (e)抓取小目标(象棋) (f)将象棋放入托盘

图 7 ABB 机械臂分拣大小目标过程

Fig.7 Process of sorting large and small objects for ABB robotic manipulator

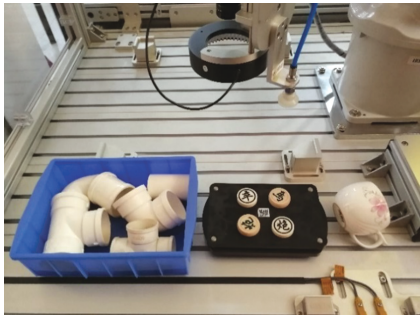


图 8 混杂目标分拣完成

Fig.8 Clutter objects sorting completion

5 结 论

1) 本文提出了一种结合全局 Kinect 深度传感器和局部手眼相机的视觉系统混合配置方法,解决了混杂场景中大小目标的位姿识别与自动分拣问题;离线阶段,基于目标的 CAD 模型生成一系列二维视图,并储存在形状模板数据库中;在线阶段,通过图像金字塔分层搜索匹配识别,找到目标的所有实例和初始位姿,经过转换和修正,计算得到目标精

确的三维位姿。

2) 本文提出的位姿估计方法与基于特征的方法^[6]和基于描述符的方法^[11]相比,在位置精度、姿态精度和识别率等方面都有明显优势,识别时间和总体分拣时间都控制在工业应用允许范围之内。

3) 本文提出的视觉系统混合配置方法未来可以广泛应用在机械臂对大小目标混杂无序放置的分拣任务中,能够明显改善机械臂自主作业的柔性和自动化程度,减少了人机交互过程。

参考文献

- [1] KAIPA K N, KANKANHALLI N A S, KUMBLA N B, et al. Addressing perception uncertainty induced failure modes in robotic bin-picking [J]. *Robotics and Computer Integrated Manufacturing*, 2016, 42: 17. DOI: 10.1016/j.rcim.2016.05.002
- [2] KRAININ M, HENRY P, REN X, et al. Manipulator and object tracking for in-hand 3D object modeling [J]. *International Journal of Robotics Research*, 2011, 30 (11): 1311. DOI: 10.1177/0278364911403178
- [3] BRATANIC B, PERNUŠ 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
- [4] 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
- [5] ZHANG H R, CAO Q X. Texture-less object detection and 6D pose estimation in RGB-D images [J]. *Robotics & Autonomous Systems*, 2017, 95: 64. DOI: 10.1016/j.robot.2017.06.003
- [6] COSTA M S, SHAPIRO L G. 3D object recognition and pose with relational indexing [J]. *Computer Vision & Image Understanding*, 2000, 79 (3): 364. DOI: 10.1006/cviu.2000.0865
- [7] 刘正琼, 万鹏, 凌琳, 等. 基于机器视觉的超视场工件识别抓取系统 [J]. *机器人*, 2018, 40 (3): 294. DOI: 10.13973/j.cnki.robot.170365.
LIU Zhengqiong, WAN Peng, LING Lin, et al. Recognition and grabbing system for workpieces exceeding the visual field based on machine vision [J]. *ROBOT*, 2018, 40 (3): 294. DOI: 10.13973/j.cnki.robot.170365
- [8] DAVID P, DEMENTHON D, DURAISWAMI R, et al. Simultaneous pose and correspondence determination using line features [C]// *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2003: 2. DOI: 10.1109/CVPR.2003.1211499
- [9] DEMENTHON D, DEMENTHON D. Object recognition in high clutter images using line features [C]// *Tenth IEEE International Conference on Computer Vision*. Beijing: IEEE Computer Society, 2005: 1581. DOI: 10.1109/ICCV.2005.173
- [10] WU K, RANASINGHE R, DISSANAYAKE G. A fast pipeline for textured object recognition in clutter using an RGB-D sensor [C]// *International Conference on Control Automation Robotics & Vision*. Piscataway: IEEE, 2014: 1650. DOI: 10.1109/ICARCV.2014.7064563
- [11] BOUAGAR S, LARABI S. Efficient descriptor for full and partial shape matching [J]. *Multimedia Tools & Applications*, 2016, 75 (6): 2989. DOI: 10.1007/s11042-014-2417-0
- [12] ZHANG J, ZHAO G, GU H. DP-PMK: An improved pyramid matching kernel for approximating correspondences in high dimensions [J]. *Neural Computing & Applications*, 2012, 21 (6): 1167. DOI: 10.1007/s00521-012-0953-y
- [13] LEPETIT V, FUA P. Keypoint recognition using randomized trees [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2006, 28 (9): 1465. DOI: 10.1109/TPAMI.2006.188
- [14] ZHANG H, EL-GAALY T, ELGAMMAL A, et al. Factorization of view-object manifolds for joint object recognition and pose estimation [J]. *Computer Vision & Image Understanding*, 2015, 139 (C): 89. DOI: 10.1016/j.cviu.2015.03.014
- [15] TEJANI A, TANG D, KOUSKOURIDAS R, et al. Latent-class hough forests for 3D object detection and pose estimation [C]// *European Conference on Computer Vision*. Cham: Springer, 2014: 462. DOI: 10.1007/978-3-319-10599-4_30
- [16] 任秉银, 魏坤, 代勇. 一种非结构环境下目标识别和 3D 位姿估计的新方法 [J]. *哈尔滨工业大学学报*, 2019, 51 (1): 38.
REN B Y, WEI K, DAI Y. A novel method of target recognition and 3D pose estimation in unstructured environment [J]. *Journal of Harbin Institute of Technology*, 2019, 51 (1): 38. DOI: 10.11918/j.issn.0367-6234.201804210
- [17] STEGER C. Occlusion, clutter, and illumination invariant object recognition [J]. *International Archives of Photogrammetry and Remote Sensing*, 2002, 34 (3/A): 345. DOI: 10.1007/3-540-45404-7_20
- [18] 克里斯蒂安·维德曼, 马库斯·乌尔里希, 卡斯滕·斯蒂格. 用于三维目标识别的系统和方法: CN101408931A [P]. 2009-04-15.
- [19] WEI K, REN B Y. A method on dynamic path planning for robotic manipulator autonomous obstacle avoidance based on an improved RRT algorithm [J]. *Sensors*, 2018, 18 (2): 571. DOI: 10.3390/s18020571

(编辑 杨 波)