

可变环境下基于位姿变换矩阵的机器人 无标定手眼协调方法

金紫凤, 潘思聪, 危 辉

(复旦大学计算机科学技术学院认知算法模型实验室, 上海 200438)

摘 要: 智能机器人不同于工业流水线上固定任务的机器人, 它们往往需要面对背景、目标物体形状、位置、姿态、尺寸的多种不确定性, 这就要求机器人能够像人一样具有极好的手眼协调能力, 能基于动态反馈临机调整自己的动作. 本文采用基于 RGB-D 相机的三维深度信息采集手段, 构建了一个用于无标定可变场景机器人手眼协调方法, 它能够利用几何关系理解三维场景, 实时跟踪并准确分割机械臂末端执行器和待抓取的物体, 在没有 3D 模型条件下计算末端执行器的姿态, 通过计算夹持器的姿态以及夹持器和目标对象在相机坐标系中的相对位移向量之后, 将该相对位移矢量从相机坐标系转换到夹持器坐标系, 估计夹持器位姿变换矩阵, 并将齐次变换矩阵输入到机械手逆运动学程序中, 求解出机械臂对应的各关节扭角, 从而驱动夹持器指向或移动到物体. 这一方法硬件资源需求量少、计算速度快、实时性好.

关键词: 手眼协调; 无标定; 三维环境感知; 计算几何; 机器人

中图分类号: TP391

文献标识码: A

文章编号: 0372-2112(2022)10-2318-11

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20201353

Uncalibrated Hand Eye Coordination Method for Robot Based on Pose Transformation Matrix in Variable Environment

JIN Zi-feng, PAN Si-cong, WEI Hui

(Laboratory of Algorithms for Cognitive Models, School of Computer Science, Fudan University, Shanghai 200438, China)

Abstract: Intelligent robots are different from the fixed task robots on the industrial assembly line. They often need to face a variety of uncertainties, such as the background, the shape, position, posture and size of the target object. This requires that the robot can have excellent hand eye coordination ability like us, and adjust its actions based on dynamic feedback. In this paper, a 3D depth information acquisition method based on RGB-D camera is used to construct a hand eye coordination method for uncalibrated variable scene robot. It can understand the 3D scene by using geometric relations, track and segment the end effector and the object to be grasped in real time, calculate the pose of the end effector without 3D model, and calculate the pose of the gripper. After the relative displacement vector of the gripper and the target object in the camera coordinate system, the relative displacement vector is transformed from the camera coordinate system to the gripper coordinate system, and the pose transformation matrix of the gripper is estimated, and the homogeneous transformation matrix is input into the inverse kinematics program of the manipulator to solve the corresponding joint torsion angles of the manipulator, so as to drive the gripper to point or move to the object. This method has the advantages of less hardware resource requirement, fast computing speed and good real-time performance.

Key words: hand-eye coordination; uncalibrated; 3D environment perception; computational geometry; robot

1 引言

机器人执行目标物体背景、位置、姿态、大小等可
变任务时对手眼协调有高度的依赖, 就像人的肢体在

可变环境中从事各种新异操作那样. 只有采用基于视
觉反馈的闭环控制(即手眼协调)才能在无需事先严格
标定的情况下适应操作环境、操作目标、位置、姿态等

的临时变化。

手眼协调可以被分为两种方式:有标定和无标定。基于标定^[1-3]的手眼协调方法是利用摄像机的内参和外参、以及机器人正逆运动学找到摄像机空间与机器人基座空间之间的对应关系,这种方法需要花费大量的时间,对其得到的参数的准确性也有很高的要求。而且当摄像机位置、焦距改变的时候,这些参数就需要被重新标定。因此,基于标定适用于工厂车间等事先可以完全预知可控制的环境。而无标定的方式更加具备灵活性,适用于非预设和动态场景中,例如利用移动的机器人采摘水果,或者在医院,图书馆等地方工作。

无标定手眼协调有着悠久的历史。Herve^[4]论证了从图像空间到机器人空间映射的可行性。Cooperstock^[5]等人利用神经网络方法控制一个不需要标定手眼关系的机器人。Su 等人^[6]提出了一种非线性视觉映射模型来替换图像雅可比变换关系,以实现无标定的手眼协调。Xiang 等人^[7]等设计了一种基于卡尔曼滤波的图像雅可比矩阵方法用来进行无标定手眼协调。

基于学习的方法用到了多个传感器^[8-11],大量训练数据,至少要经过几个小时的训练过程,并且不能迅速移植到一个新的环境里。因此,这一方法并不适用于非预设和开放环境的任务。基于雅可比矩阵的方法一般需要在夹持器工作前,预先人工给出 n 步线性独立的试探运动,记录图像特征的变化来估计实验需要的雅可比矩阵,为了辨识夹持器,一般需要在夹持器上做标记,通过不断缩小夹持器和目标物体在图像上的差距来完成跟踪任务,因而自动化程度不高。

形状是重要的图像特征,利用二维 RGB 图像上的轮廓信息可以有效实现无标定手眼协调^[12]。为实现自动化程度较高的无标定手眼协调,本文提出的方法采用三维空间中时间上连续的 RGB-D 图像上的形状特征对物体和夹持器进行姿势位置的跟踪。则可在计算出相机坐标系中夹持器的姿态以及夹持器到目标对象的相对位移向量之后,计算出旋转矩阵或平移向量,并将变换矩阵输入到机械臂逆运动学程序中,从而驱动机械臂。该方法使机器人快速准确地移动到目标物体的位置,反应灵敏,且应对非预设环境。

本文方法需要计算出夹持器的姿势,现有的基于 RGB 的位姿估计方法包括 Line-2D^[13], PWP3D^[14] 和 SSD-6D^[15]。随着廉价深度传感器的普及,基于 RGB-D 的方法应运而生,例如经典的 Linemod^[16] 和 PoseCNN^[17]。然而,现有的方法存在以下缺点:(1)其位姿估计方法需要对精确物体的三维原型有预先的了

解;(2)以 PoseCNN 为代表的深度学习方法需要大量的离线训练和在线检测计算,必须提供 GPU 等硬件设备来训练以快速预测姿势位置。对于不依赖深度学习的姿态跟踪,在估计姿态的前一步就是分割出夹持器的点云。在本实验中,采用 OBB 最小包围盒的方法估计其姿态,这一方法适应于识别机械臂末端夹持器。

2 实验环境

图 1 显示了本研究中搭建的机器人操作平台及安装的硬件设备。本文将 RGB-D 深度相机和双目相机架设在机械臂正前方上侧 1.0 m 左右的位置,具有六自由度的 SD700E 机械臂安装于桌面上,夹持器安装在机器人机械手末端。图中桌面上的小犀牛是待抓握的对象。

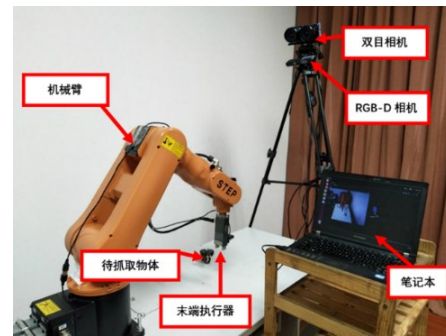


图 1 SD700E 机器人工作台

在本文实验中,仅仅使用一台普通笔记本电脑完成所有运算,在笔记本上安装深度相机驱动后,相机的图像通过 USB 接口可以实时传输到笔记本电脑上。

图 2 中的相机坐标系,是以相机的光心为原点 O , z 轴正向指向摄影机的正前方,由右手法则建立的坐标系,其中 y 轴垂直地面。光学中心 O 是相机的光学中心。

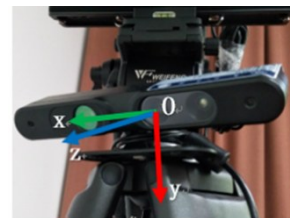


图 2 RGB-D 相机及其坐标系

由针孔相机模型可知,像素坐标 (u, v) 和相机坐标 (x, y, z) 之间的关系如下:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \frac{f_u x}{z} + c_x \\ \frac{f_v y}{z} + c_y \end{bmatrix} \quad (1)$$

其中, f_u, f_v 是 x 轴和 y 轴方向上的每单位长度的像素个数;而 c_x, c_y 则是 x 轴和 y 轴方向上的光心坐标。

给定一个像素的深度 z , 它的空间坐标为

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{z(u-c_x)}{f_u} \\ \frac{z(v-c_y)}{f_v} \\ z \end{bmatrix} \quad (2)$$

深度相机测量了深度 z , 所以 x 和 y 可以被推断出来从而重建点云, 如图 3 所示.



图3 三维重建SD700E实验环境的点云

3 点云分割

3.1 机械臂夹持器分割

通常来说, 机械臂的夹持器生产厂商都会提供其 3D 模型. 因此, 可以利用一些基于 3D 模型的位姿识别方法来获取夹持器第一帧的位姿数据, 例如 DenseFusion^[18], CosyPose^[19], Hybrid-DL-PointPairs^[20]等. 本文采用了基于 PPF^[21,22]的方法在开机的时候获取初始位姿. 此外, 为了在将来的图像中更准确地识别机械臂的夹持器, 同时也为了适用于厂商未提供 3D 模型的情况, 本文也提供了较为简便的人机交互环节, 在开机时操作人员标注夹持器位置的大致范围(即一个包围盒), 如所示的橙色方框内. 为了分割夹持器, 拍摄没有夹持器的图片并将其转换为灰度图像, 如图 4(b)所示. 然后拍摄带有夹持器的图像并将其转换为灰度图像, 如图 4(a)所示. 通过背景减法确定两者之间的差异, 并对阈值 THRESH 进行二值化以获得图 4(c)中的结果, 通过二值化提取夹持器并突出显示, 其中 i 和 j 分别表示点的水平和垂直坐标:

$$c(i,j) = \begin{cases} 0, & a(i,j) - b(i,j) < \text{THRESH} \\ 255, & a(i,j) - b(i,j) > \text{THRESH} \end{cases} \quad (3)$$



(a) 原始图片 (b) 背景图片 (c) 二值化差异图

图4 背景减法示意图

为了连续引导机械手的运动, 对夹持器在第一帧中标出边界框后进行跟踪. 在本文实验中, 使用经典的基于核相关滤波器 KCF^[23]的目标跟踪方法来跟踪夹持器.

KCF 算法存在一些缺点, 比如当目标的位置靠近边缘的时候, 损失了部分目标的信息, 不能自适应目标尺度变化等. 另有一些目标跟踪的算法虽然在一定程度上可以克服这些缺点, 但是却不能达到实时性的要求. 而且在本文的实验环境中, 机械臂运动速度稳定, 目标位置与上一帧的位置相差不大, 目标始终在跟踪框内. 机械臂距离摄像头的距离大致不变, 所以机械臂夹持器的大小不会发生显著变化, 且不存在遮挡现象, 从而避免了边界效应和过大的尺度变化.

由于机械臂夹持器的分割轮廓往往不够精确, 可能会在夹持器周围带有背景上的点, 所以需要通过统计滤波器移除离群点.

3.2 机械臂夹持姿态估计

对于分割而得的夹持器点云, 接着要对滤波后的点云进行姿态的计算. 用一个最小的外接六面体包围滤波后的点云, 并用该六面体相互正交的 3 个方向来表示点云的姿态方向. OBB 全称为 Oriented Bounding Box, 是计算某一物体最小包围盒的一种方法. 夹持器的点云被一个最小外切六面体包围, 其姿态方向由六面体的 3 个正交方向表示. OBB^[24]是一种计算对象最小边界框的方法. 求解 OBB, 就是找到 3 个线性独立的轴, 使点云在这些轴上的投影尽可能分离, 并使投影样本点的方差最大化. 通过主成分分析可以找到 3 个特征向量作为位姿估计的结果. 对于图 5 列出的规则形状与不规则形状的夹持器(包含一个刚体连接的六面体), 这样一种基于包围盒的算法是可行的, 效果如图 6 所示.

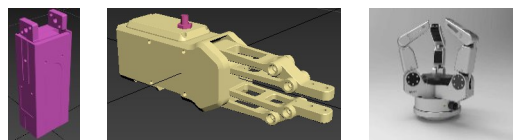


图5 各类夹持器

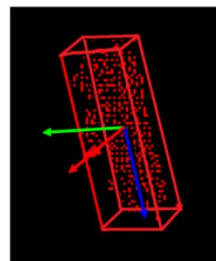


图6 夹持器 3 个轴的图示

最简单的包围盒方法有助于找到夹持器的姿态,虽然不完全精确到每个空间点的位置,但对于本文实验来说已经足够验证无标定手眼协调方法的有效性.

同样,在对目标物体的点云进行分割后,利用该方法可以计算出目标的姿态.因为在某些特定的任务中,可能需要处理目标物体的不同姿态来引导机械手的运动.

3.3 点云分割

3.3.1 基于RANSAC的桌面拟合

为了减少无关对象的干扰,本文希望能仅仅保留桌面上的物体,且为了能对桌面上的物体进行点云分割,首先应该过滤掉属于桌面的点云.所以先利用RANSAC算法[25]对桌面进行拟合和抽取,再将投影在桌面范围内的部分留下.

3.3.2 无关点云的去除

所有点都投影到桌面的平面上,只留下落在桌面凸包内的点并删除其他点.

具体处理步骤如下.首先对机器人手臂的RGB图像进行二值化,通过形态学运算,得到桌面的凸包.由此可以计算桌面平面的一般方程中的常数 A, B, C, D ,即

$$Ax + By + Cz + D = 0 \quad (4)$$

抽掉桌面后,点云中的其他点投影到桌面所在的平面上,从点云中的某点 (x_0, y_0, z_0) 作垂线到桌面所在的平面上.投影点为 (x_p, y_p, z_p) ,形式如下:

$$\frac{x_0 - x_p}{A} = \frac{y_0 - y_p}{B} = \frac{z_0 - z_p}{C} \quad (5)$$

因此可以计算投影点的坐标,即

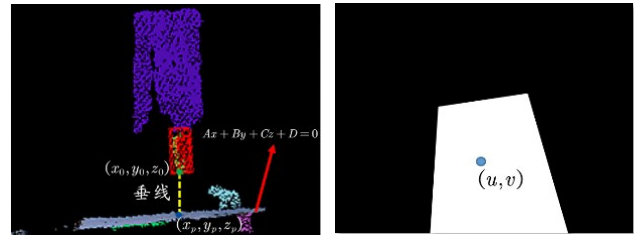
$$\begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix} = \begin{bmatrix} \frac{(B^2 + C^2)x_0 - A(By_0 + Cz_0 + D)}{A^2 + B^2 + C^2} \\ \frac{(B^2 + C^2)y_0 - A(Bx_0 + Cz_0 + D)}{A^2 + B^2 + C^2} \\ \frac{(B^2 + C^2)z_0 - A(Bx_0 + Cy_0 + D)}{A^2 + B^2 + C^2} \end{bmatrix} \quad (6)$$

根据式(7)可以计算出二值图像上投影点的二维坐标,进而得到像素的颜色,判断是否落入凸包,即

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \frac{f_u x_p}{z_p} + c_x \\ \frac{f_v y_p}{z_p} + c_y \end{bmatrix} \quad (7)$$

如图7所示,如果二值图像中的像素 (u, v) 的颜

色是白色,则它位于桌面上方,因此保留它;但如果颜色为黑色,则超出了桌面的范围,因此将其移除.



(a) 点云的投影到桌面的过程 (b) 二维图像中桌面的凸包

图7 无关点云去除示意图

3.4 各类形状对象分割

从点云中移除桌面,桌面上的对象将被悬空.因而可使用欧几里德聚类算法[2]分割点云.它直接使用距离来分隔非内聚对象.移除桌面后,组件是独立的.于是无论对象的形状如何,都可找到一个OBB来包围该对象,随后通过计算3个主轴来获取其位姿,如图8所示.

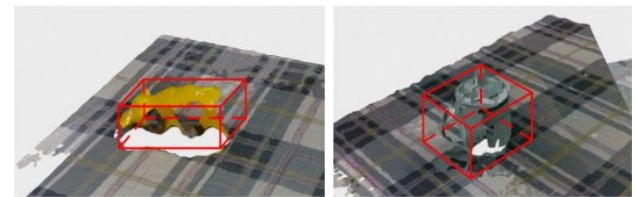


图8 不同形状物体的包围盒

本文使用了一台带有 intel i5-3320M CPU 和 4 GB RAM 的笔记本电脑来进行点云分割的实验,这其中还包括了图像数据的通信传输,无需任何其他计算强化硬件的辅助.在不同的深度相机安置位置、距离,不同的桌面上物体的环境下,6组实验均得到了良好的效果,如图9所示.由此可以看出,本文方法对传感器位置和距离变化具有很好的耐受性,对物体的分辨能力也很好.后续可以计算出以边界盒为例的目标物体的3个轴来指导后续路径规划实验.

4 机器人手眼协调运动规划

4.1 标记三个轴的方向

在上面的视觉实验中,找到夹持器的3个轴,并标记这3个轴的正方向.

夹持器可以由图10表示.将所述坐标系的原点放在夹持器的中心,该中心由向量 p 表示.描述夹持器3个方向的3个单位向量的方向如下: z 轴方向向量是在

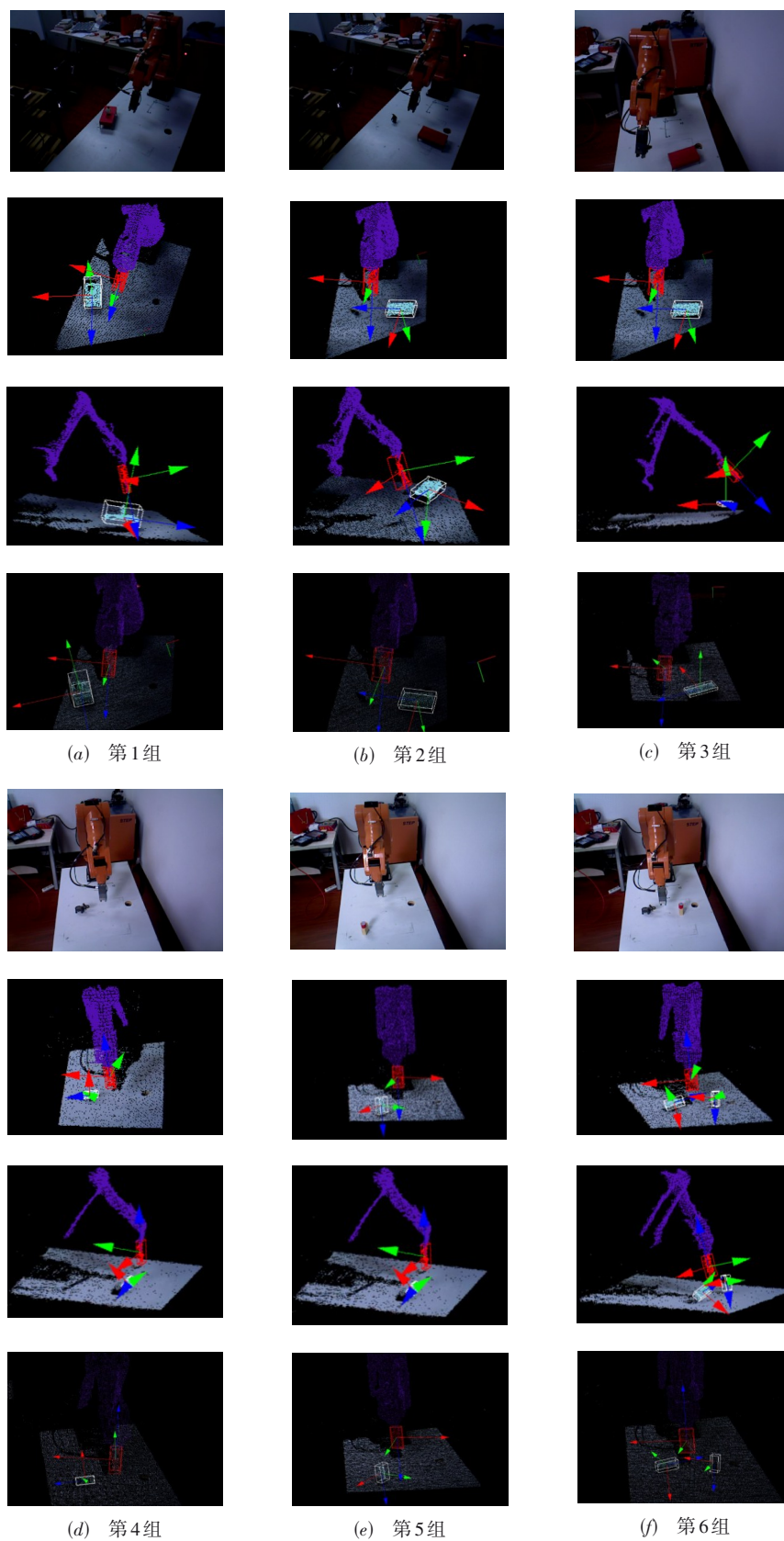


图9 改变RGB-D相机与桌面物体的位置和距离后的6组实验结果
注:从上到下分别是点云的原始图像、前视图、侧立面图和放大图像

夹持器朝向方向上,即所谓的接近向量 \mathbf{a} ,就像图 10 中的蓝色箭头一样;y 轴方向向量是从一个指尖到另一个指尖,穿过夹持器的宽侧被称为方向向量 \mathbf{o} ,如图 10 中的绿色箭头所示;最后一个向量称为法向量 \mathbf{n} ,如图 11 中的红色箭头所示,它与向量 \mathbf{a} 和 \mathbf{o} 一起构成右手向量集.

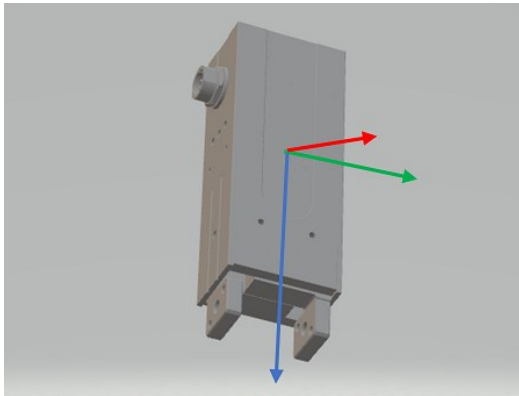


图 10 夹持器的坐标系

利用空间位置关系和右手定则确定夹持器姿态方向,计算出 3 个轴的单位矢量.

综上所述,已经计算了夹持器的 3 个轴及其方向,3 个轴在相机坐标系中,它形成了 3×3 的矩阵,表示如下:

$$\mathbf{R}_0 = [\mathbf{n} \quad \mathbf{o} \quad \mathbf{a}] = \begin{bmatrix} x_n & x_o & x_a \\ y_n & y_o & y_a \\ z_n & z_o & z_a \end{bmatrix} \quad (8)$$

4.2 求解期望指向向量

如图 11 所示,在相机坐标系中夹持器中心的坐标 p 为 (x_p, y_p, z_p) ,夹持器期望指向的方向向量,即从夹持器中心到物体包围盒中心向量 \mathbf{v}_c 为

$$\mathbf{v}_c = (x_t - x_p, y_t - y_p, z_t - z_p) \quad (9)$$

在夹持器坐标系中,向量 \mathbf{v}_c 可以被 \mathbf{R}_0 的 3 个线性

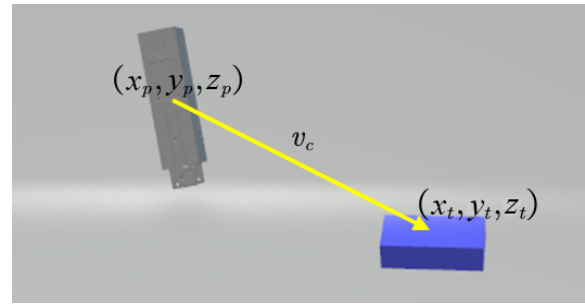


图 11 在相机坐标系中的期望指向的向量示意图

无关的向量表示,线性组合的系数即为在夹持器笛卡尔坐标系中的坐标,即

$$\mathbf{v}_g = \mathbf{R}_0^{-1} \mathbf{v}_c \quad (10)$$

4.3 指向目标物体

为了使夹持器指向物体,即接近向量 \mathbf{a} 的方向指向物体,原先的接近向量 \mathbf{a}_g 为

$$\mathbf{a}_g = (0, 0, 1) \quad (11)$$

使夹持器指向物体的中心,从向量 \mathbf{a}_g 到 \mathbf{v}_g 的旋转向量可以通过两者之间的叉乘计算出来. k 是旋转轴, θ 是旋转角,即

$$k = \frac{\mathbf{a}_g \times \mathbf{v}_g}{|\mathbf{a}_g \times \mathbf{v}_g|} \quad (12)$$

$$\theta = |\mathbf{a}_g \times \mathbf{v}_g| \quad (13)$$

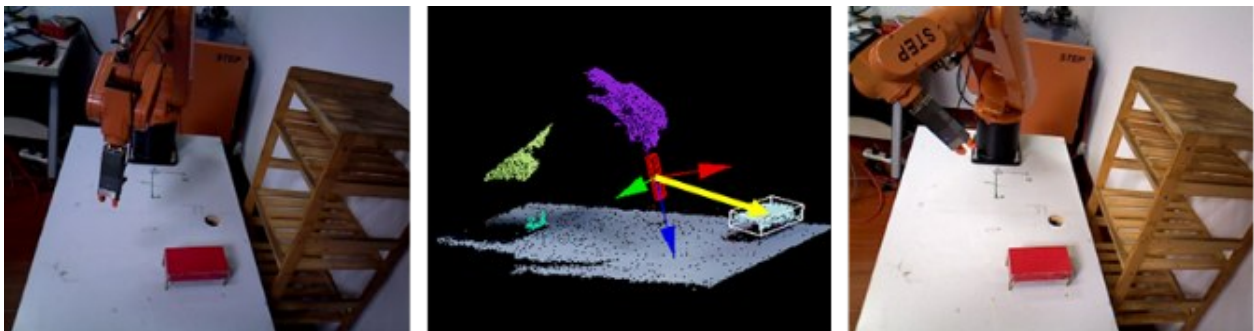
已知旋转轴和相应的旋转角度,则旋转矩阵可以根据罗德里格斯公式求出,即

$$\mathbf{R} = \cos\theta \mathbf{I} + (1 - \cos\theta) \mathbf{k} \mathbf{k}^T + \sin\theta \mathbf{k} \quad (14)$$

于是用于指导机械臂运动的运动学矩阵 ${}^0\mathbf{T}$ 可以得到,且输入到机械臂的控制程序中,即

$${}^0\mathbf{T} = \begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix} \quad (15)$$

随后的实验证明此推理是正确的,把矩阵输入程序求得反解^[26]确实可以引导机械手指向目标,如图 12 所示.



(a) 旋转前的照片

(b) 对应分割的点云以及估计的夹持器姿态,黄色箭头代表了需要指向的方向

(c) 旋转后的结果

图 12 朝向对象旋转的操纵器夹持器示意图

4.4 趋近目标物体

为了使机械手的手指接近物体,保持夹持器的姿

态不变,并使夹持器移动一段距离,即从指尖到物体的矢量,如图13所示例子的黄色向量.

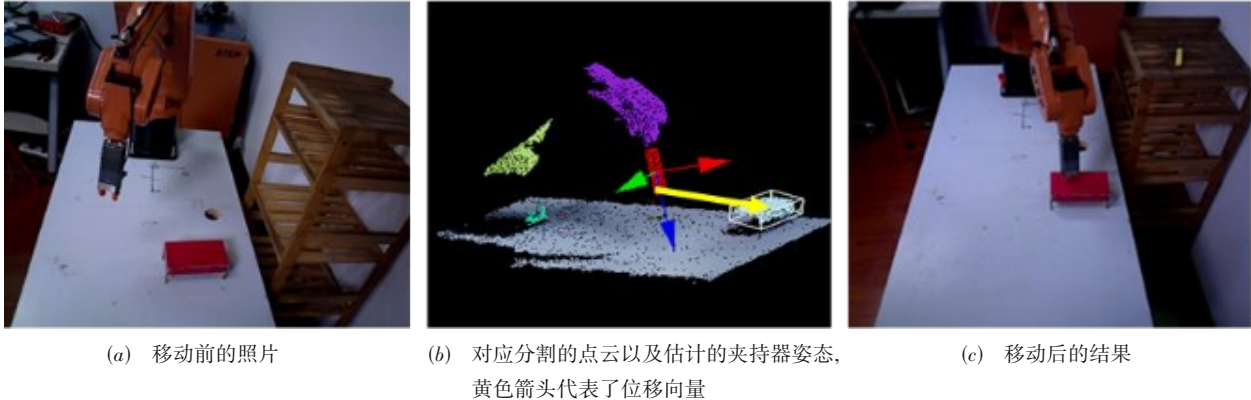


图13 利用位姿变换矩阵使机械手夹持器趋近目标物体的实验图例

如前文所述,目标物体在夹持器坐标系中的坐标为 \mathbf{v}_g ,正如图14中的黄色箭头所示.从包围盒的中心到指尖的向量为 \mathbf{a}_f 相当于 $l \times \mathbf{a}$, l 是从夹持器的中心到指尖的长度,如图13中橘色的向量所示.所以从指尖到物体的向量为 $\mathbf{v}_g - \mathbf{a}_f$,如图13中的黑色箭头.随后得到指导机械臂的变化矩阵 0T ,即

$${}^0T = \begin{bmatrix} \mathbf{I} & \mathbf{v}_g - \mathbf{a}_f \\ \mathbf{0}^T & 1 \end{bmatrix} \quad (16)$$

其中, \mathbf{I} 代表单位矩阵.

根据上述计算矩阵,将这个齐次变换矩阵代入机械臂的运动学方程中,即可以求解出各关节的扭转角.实验表明,该方法是有用的.

在计算出变换矩阵并输入到机械臂逆解程序中后,夹持器在不改变姿态的情况下移动到目标位置.从后续统计分析的结果得到,运动平均误差为 5.389 93 mm.图14是一个例子,从中可以看出,本文的识别算法能准确估计手爪与物体的位姿.

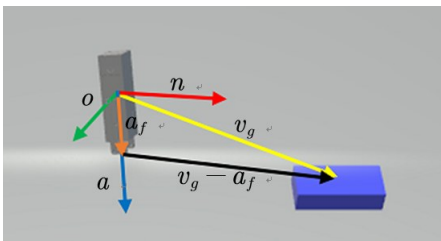


图14 夹持器坐标系与指尖到目标物体的向量

5 误差分析

5.1 引入球坐标系

根据上述旋转实验的描述,从机械臂夹持器中心指向目标物体的向量在夹持器坐标系中的表示为

$$\mathbf{v}_g = \mathbf{R}_0^{-1} \mathbf{v}_c = \mathbf{R}_0^T \mathbf{v}_c = \begin{bmatrix} \mathbf{n}^T \mathbf{v}_c \\ \mathbf{o}^T \mathbf{v}_c \\ \mathbf{a}^T \mathbf{v}_c \end{bmatrix} \quad (17)$$

根据罗德里格斯公式的要求,应计算两个向量的叉积,令

$$\mathbf{f} = \mathbf{a}_g \times \mathbf{v}_g = \begin{bmatrix} -\mathbf{o}^T \mathbf{v}_c \\ \mathbf{n}^T \mathbf{v}_c \\ 0 \end{bmatrix} = \begin{bmatrix} -|\mathbf{v}_c| |\mathbf{o}^T \frac{\mathbf{v}_c}{|\mathbf{v}_c}| \\ |\mathbf{v}_c| |\mathbf{n}^T \frac{\mathbf{v}_c}{|\mathbf{v}_c}| \\ 0 \end{bmatrix} \quad (18)$$

$\mathbf{n}, \mathbf{o}, \mathbf{v}_c / |\mathbf{v}_c|$ 都是单位向量,它们可以被球坐标系表示.令 $\theta_n, \varphi_n, \theta_o, \varphi_o, \theta_{v_c}, \varphi_{v_c}$ 依次作为 $\mathbf{n}, \mathbf{o}, \mathbf{v}_c$ 在球坐标系中的仰角和方位角.相应地,求解坐标系中的仰角 θ 和方位角 φ ,如图15所示.

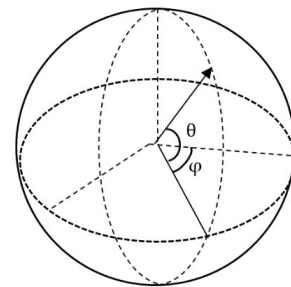


图15 球坐标系中的仰角 θ 和方位角 φ

因此 $\mathbf{f} = (f_0, f_1, f_2)^T$ 可以由式(19)表示:

$$\mathbf{f} = |\mathbf{v}_c| \begin{bmatrix} -\cos\theta_o \cos\varphi_o & -\cos\theta_o \sin\varphi_o & -\sin\theta_o \\ \cos\theta_n \cos\varphi_n & \cos\theta_n \sin\varphi_n & \sin\theta_n \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \cos\theta_{v_c} \cos\varphi_{v_c} \\ \cos\theta_{v_c} \sin\varphi_{v_c} \\ \sin\theta_{v_c} \end{bmatrix} \quad (19)$$

5.2 误差传播

令 $z = z(x_1, x_2, \dots, x_n)$, 其中 x_1, x_2, \dots, x_n 的均方差 $m_i (i=1, 2, \dots, n)$. 则根据全微分方程计算总体误差如下:

$$m_z^2 = \left(\frac{\partial z}{\partial x_1}\right)^2 m_1^2 + \left(\frac{\partial z}{\partial x_2}\right)^2 m_2^2 + \dots + \left(\frac{\partial z}{\partial x_n}\right)^2 m_n^2 \quad (20)$$

其中, m_z^2 是观察方程的方差, $\partial z / \partial x_i$ 是函数 z 对 x_i 的偏导数. 方差越小, 收敛越快. 假设两次计算结果分别为 f, f' , 两者的误差为 Δf , 即

$$\Delta f = f - f' \quad (21)$$

Δf 越接近 0, 结果越可信.

$$m_{\Delta f}^2 = m_f^2 - m_{f'}^2 \quad (22)$$

求函数 f 对 θ_n, φ_n 的偏导数如下:

$$\frac{\partial f}{\partial \theta_n} = |v_c| \left(\cos \theta_n \sin \theta_{v_c} - \sin \theta_n \cos \theta_{v_c} \left(\cos(\varphi_n - \varphi_{v_c}) \right) \right) \quad (23)$$

$$\frac{\partial f}{\partial \varphi_n} = |v_c| \left(\cos \theta_n \cos \theta_{v_c} \sin(\varphi_{v_c} - \varphi_n) \right) \quad (24)$$

其中, $\partial f / \partial \theta_n, \partial f / \partial \varphi_n$ 为姿势估计误差传递的系数. 可以看到最终的误差与夹持器距目标物体的距离 v_c 有关, 与夹持器的 n 方向角度, 以及夹持器到目标物体的方向矢量的角度有关.

因为 f_0, f_1 有着相同的地位, n, o 也有着相同的地位, 因此 f_1 和 n 被用在公式中. 为了使结果收敛, 使结果的误差最小, 因此, 使

$$\frac{\partial f_1}{\partial \theta_n} \rightarrow 0, \frac{\partial f_1}{\partial \varphi_n} \rightarrow 0 \quad (25)$$

而 $\theta_n, \varphi_n, \theta_{v_c}, \varphi_{v_c}$ 越接近式(26)和(27)的要求, 旋转的误差越小, 即

$$\theta_n = \arctan \frac{\tan \theta_{v_c}}{\cos(\varphi_n - \varphi_{v_c})} \quad (26)$$

$$\varphi_{v_c} = \varphi_n \text{ or } \theta_n = \frac{\pi}{2} \text{ or } \theta_{v_c} = \frac{\pi}{2} \quad (27)$$

同样地, 趋近实验中的平移向量 t 为

$$t = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} = v_g - l \cdot a = \begin{bmatrix} n^T v_c \\ o^T v_c \\ a^T v_c \end{bmatrix} - l \cdot a \quad (28)$$

其中, t_1, t_2, t_3 的地位相同, n, o 的地位也与 a 相同, 所以在里仅考虑 t_3 , 即

$$t_3 = |v_c| \left(\cos \theta_a \cos \varphi_a \cos \theta_{v_c} \cos \varphi_{v_c} + \cos \theta_a \sin \varphi_a \cos \theta_{v_c} \sin \varphi_{v_c} + \sin \theta_a \sin \theta_{v_c} - l \sin \theta_a \right) \quad (29)$$

t_3 的方差为

$$m_{t_3}^2 = \left(\frac{\partial t_3}{\partial \theta_a}\right)^2 m_{\theta_a}^2 + \left(\frac{\partial t_3}{\partial \varphi_a}\right)^2 m_{\varphi_a}^2 \quad (30)$$

其中,

$$\frac{\partial t_3}{\partial \theta_a} = |v_c| \left(-\sin \theta_a \cos \theta_{v_c} \cos(\varphi_a - \varphi_{v_c}) + \cos \theta_a (\sin \theta_{v_c} - l) \right) \quad (31)$$

$$\frac{\partial t_3}{\partial \varphi_a} = |v_c| \left(-\cos \theta_a \cos \theta_{v_c} \sin(\varphi_{v_c} - \varphi_a) \right) \quad (32)$$

为了使结果收敛, 要使误差最小, 所以, 令

$$\frac{\partial t_3}{\partial \theta_a} \rightarrow 0, \frac{\partial t_3}{\partial \varphi_a} \rightarrow 0 \quad (33)$$

其中, θ_a, φ_a 越接近式(34)和(35)的要求, 平移的误差越小, 即

$$\theta_a = \arctan \frac{\sin \theta_{v_c} - l}{\cos \theta_{v_c} \cos(\varphi_a - \varphi_{v_c})} \quad (34)$$

$$\varphi_a = \varphi_{v_c} \quad (35)$$

5.3 目标物体与夹持器距离过长的影响

从上述推导过程来看, 值得注意的是, 由于偏微分方程中总是出现一个因子 $|v_c|$, 它代表目标物体与夹持器之间的距离, 因此, 当夹持器距离目标物体太远时, $|v_c|$ 过大时, 均方误差较大, 系统可能陷入振荡和不稳定状态.

当夹持器与被测物体之间的距离过大时, 为了补偿误差, 将夹持器移至目标附近位置, 进行指向实验.

5.4 手眼协调误差的统计实验图

根据前文的理论分析, 本文在实验视频的每一帧中统计了机械臂指向目标物体的误差数据.

根据式(34)和(35), 平移误差定义为

$$\left| \theta_a - \arctan \frac{\sin \theta_{v_c} - l}{\cos \theta_{v_c} \cos(\varphi_a - \varphi_{v_c})} \right| + |\varphi_a - \varphi_{v_c}| \quad (36)$$

根据式(26)和(27), 旋转误差定义为

$$\left| \theta_n - \arctan \frac{\tan \theta_{v_c}}{\cos(\varphi_n - \varphi_{v_c})} \right| + |\varphi_n - \varphi_{v_c}| \quad (37)$$

图 16 和图 17 展示了平移误差统计图, 图 18 和图 19 给出了旋转误差统计图. 本文定义的误差评估指标综合分析了影响误差的几个要素并对其求得 1 范数, 因而指标越小表示平移与旋转误差越小. 总体而言, 从图 16 和图 18 中可以看出, 平移与旋转误差并不大, $2^\circ \sim 3^\circ$ 的偏差. 换算为实际尺度下, 质心误差为 4~6 mm. 在满足实时性的要求下, 这样的手眼协调精度是可接受的. 而

图 17 和图 19 则展示了误差分布的直方图,直方图的分布接近正态分布,因此实验中更多的是随机误差.

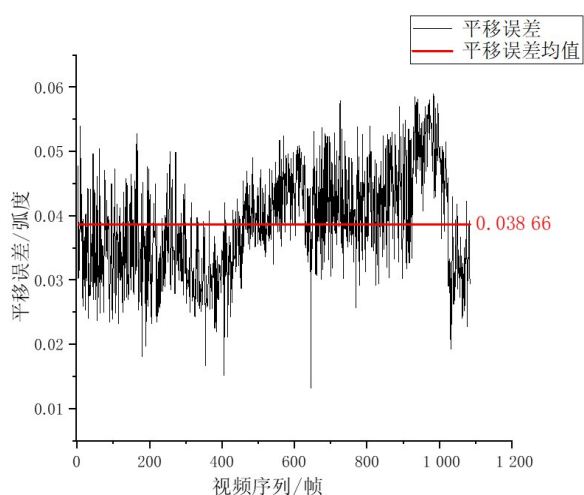


图 16 平移误差图

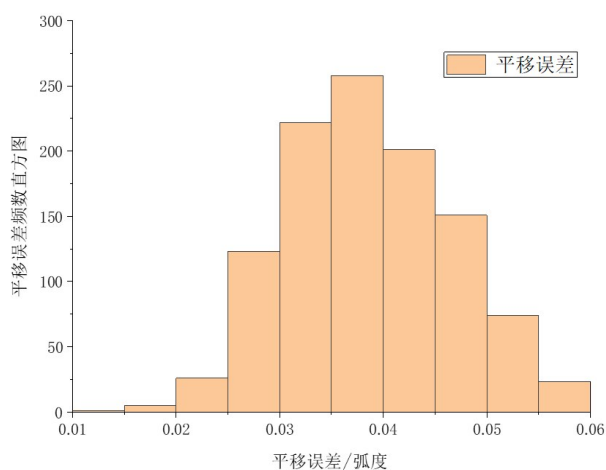


图 17 平移误差直方图

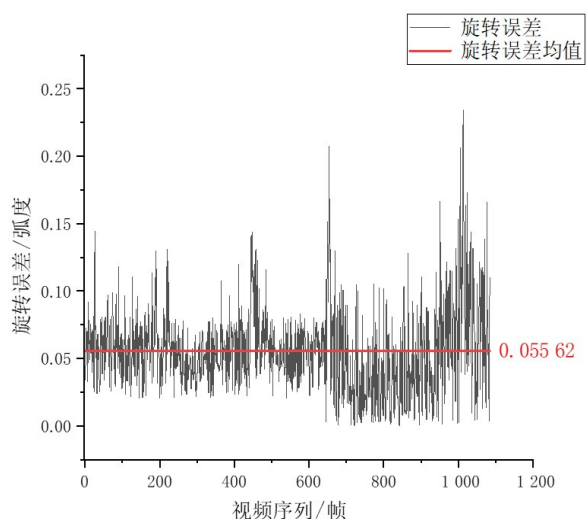


图 18 旋转误差图

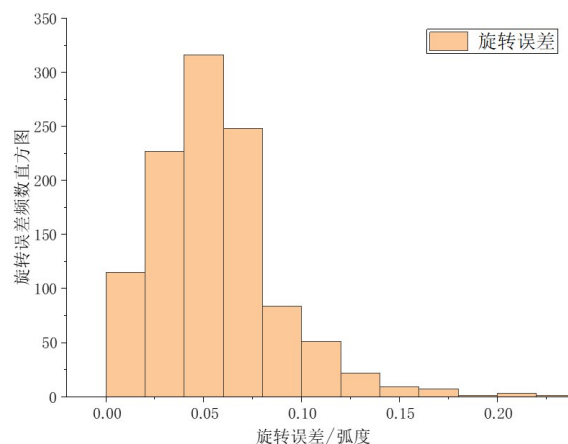


图 19 旋转误差直方图

6 对比实验

本文所述无标定手眼协调方法直接基于点云驱动,方法核心在于点云分割和末端执行器的姿态估计.

一般有纹理的物体判别其姿态都是通过图 18 旋转误差直方图,例如以 sift, surf 等方式寻找关键点,然后关键点匹配来求解旋转矩阵和平移矩阵.

无纹理对象的特征点较少,它们的外观通常由它们的投影轮廓决定. Linemod 方法采用彩色图像的边缘的梯度信息结合物体表面的法线特征作为模板匹配的依据,可以应对无纹理的情况,不需要冗长的训练时间.

使用 Linemod 方法在本文的实验里进行姿态匹配的时候,存在以下问题:

- (1) 匹配不稳定,易倾斜,如图 20 所示;
- (2) 手爪转到侧面,容易仅仅匹配到手爪的窄侧面,如图 21 所示;
- (3) 手指部分发生反光的情况下不能整体匹配,如图 22 所示;
- (4) 将正面匹配到侧面,如图 23 所示.

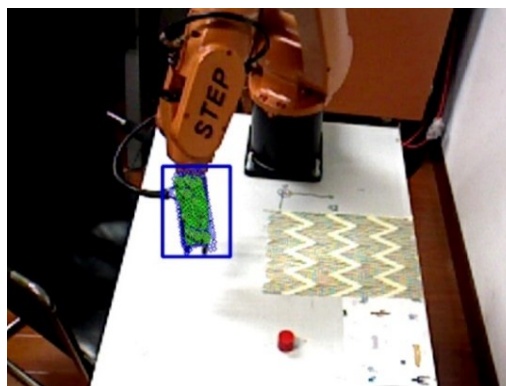


图 20 匹配不稳定示意图

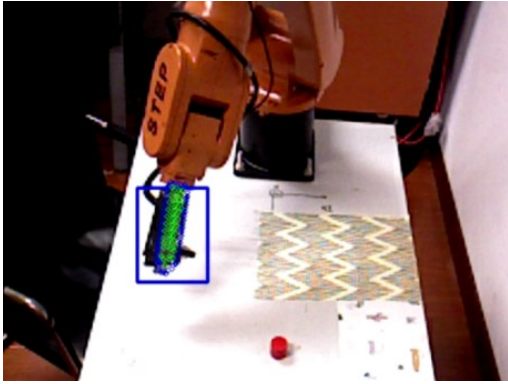


图 21 手爪转到侧面示意图

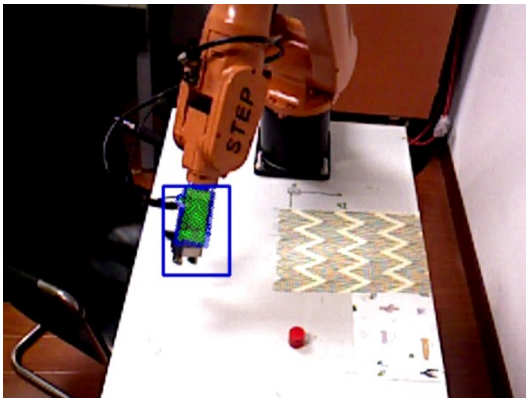


图 22 不能整体匹配示意图

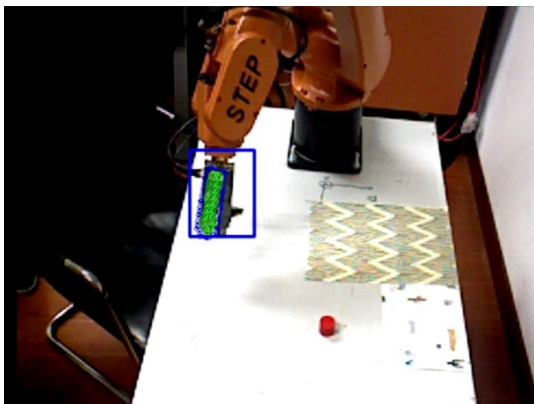


图 23 正面匹配到侧面示意图

7 结论

本文提出了一种有效的无标定手眼协调场景分割方法和简单的机械手驱动方法,即在相机坐标系下,对机器人手臂抓取物体的场景点云进行分割,并对夹持器的姿态进行估计.通过求解夹持器在相机坐标系中的位姿变换方程,使机械臂能够指向或接近目标,从而达到无标定手眼协调的效果.

可变环境下难以做到事先样本采集和训练,而本文提出的机械手驱动方法不需要大量训练数据、耗时的训练时间、复杂的手眼标定过程或者估计雅可比矩

阵前的 n 步试探运动,即可控制机械臂转向或者移动到物体.所提出的一整套方案,包括夹持器姿态的估计、点云的分割,以及机械臂转向和接近的方法,不需要高性能的硬件设备,运行速度快,可用于非预设、可变场景的理解,不依赖 3D 模型,不依赖 GPU 等硬件设备.

参考文献

- [1] ZHANG Z. A flexible new technique for camera calibration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11): 1330-1334.
- [2] ZHANG Y, QIU Z, ZHANG X. A simultaneous optimization method of calibration and measurement for a typical hand-eye positioning system[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 70: 1-11.
- [3] WANG H, FAN X, LU X. Application of a hand-eye self-calibration technique in robot vision[C]//Proceedings of the 2013 25th Chinese Control and Decision Conference (CCDC). Taiyuan: IEEE, 2013: 3765-3769.
- [4] HERVÉ J Y, SHARMA R, CUCKA P. Toward robust vision-based control: Hand/eye coordination without calibration[C]//Proceedings of the 1991 IEEE International Symposium on Intelligent Control. Arlington: IEEE, 1991: 457-462.
- [5] COOPERSTOCK J R, MILIOS E E. Self-supervised learning for docking and target reaching[J]. Robotics and Autonomous Systems, 1993, 11(3-4): 243-260.
- [6] SU J B, PAN Q L, XI Y G. Dynamic coordination of uncalibrated hand/eye robotic system based on neural network [J]. Journal of Systems Engineering and Electronics, 2001, 12(3): 45-50.
- [7] XIANG Z, SU J, MA Z. Uncalibrated hand-eye coordination based HRI on humanoid robot[C]//Proceedings of the 33rd Chinese Control Conference. Nanjing: IEEE, 2014: 8438-8443.
- [8] ZENG A, SONG S, YU K T, et al. Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching[C]//Proceedings of the 2018 IEEE International Conference on Robotics and Automation(ICRA). Brisbane: IEEE, 2018: 3750-3757.
- [9] ZENG A, YU K T, SONG S, et al. Multi-view self-supervised deep learning for 6d pose estimation in the amazon picking challenge[C]//Proceedings of the 2017 IEEE International Conference on Robotics and Automation(ICRA). Singapore: IEEE, 2017: 1386-1383.
- [10] RUTKOWSKI S, ADAMCZYK M, PASTUŁA A, et al. Training using a commercial immersive virtual reality

- system on hand-eye coordination and reaction time in young musicians: A pilot study[J]. International Journal of Environmental Research and Public Health, 2021, 18(3): 1297.
- [11] BOYD L C, POPOVIC V, SIEBERT J P. Deep reinforcement learning control of hand-eye coordination with a software retina[C]//Proceedings of the 2020 International Joint Conference on Neural Networks(IJCNN). Shenzhen: IEEE, 2020: 1-8.
- [12] WEI A H, CHEN B Y. Robotic object recognition and grasping with a natural background[J]. International Journal of Advanced Robotic Systems, 2020, 17(2): 1729881420921102.
- [13] HINTERSTOISSER S, HOLZER S, CAGNIART C, et al. Multimodal templates for real-time detection of texture-less objects in heavily cluttered scenes[C]//Proceedings of the 2011 IEEE International Conference on Computer Vision(ICCV). Barcelona: IEEE, 2011: 858-865.
- [14] PRISACARIU V A, REID I D. PWP3D: Real-time segmentation and tracking of 3D objects[J]. International Journal of Computer Vision, 2012, 98(3): 335-354.
- [15] KEHL W, MANHARDT F, TOMBARI F, et al. Ssd-6d: Making rgb-based 3d detection and 6d pose estimation great again[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 1521-1529.
- [16] HINTERSTOISSER S, CAGNIART C, ILIC S, et al. Gradient response maps for real-time detection of textureless objects[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 34(5): 876-888.
- [17] XIANG Y, SCHMIDT T, NARAYANAN V, et al. Posecnn: A Convolutional Neural Network for 6d Object Pose Estimation in Cluttered Scenes[EB/OL]. (2017-11-01)[2020-11-30]. <https://arxiv.org/abs/1711.00199>.
- [18] WANG C, XU D, ZHU Y, et al. Densefusion: 6d object pose estimation by iterative dense fusion[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE/CVF, 2019: 3343-3352.
- [19] LABBÉ Y, CARPENTIER J, AUBRY M, et al. CosyPose: Consistent multi-view multi-object 6D pose estimation[C]//Proceedings of the 2020 European Conference on Computer Vision. Glasgow: Springer, 2020: 574-591.
- [20] FU C Y, SHVETS M, BERG A C. RetinaMask: Learning to Predict Masks Improves State-of-the-Art Single-Shot Detection for Free[EB/OL]. (2019-01-10) [2020-11-30]. <https://arxiv.org/abs/1901.03353>.
- [21] KIFORENKO L, DROST B, TOMBARI F, et al. A performance evaluation of point pair features[J]. Computer Vision and Image Understanding, 2018, 166: 66-80.
- [22] HINTERSTOISSER S, LEPETIT V, RAJKUMAR N, et al. Going further with point pair features[C]//Proceedings of the 2016 European Conference on Computer Vision. Amsterdam: Springer, 2016: 834-848.
- [23] HENRIQUES J F, CASEIRO R, MARTINS P, et al. High-speed tracking with kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 37(3): 583-596.
- [24] LIVERANI A, AMATI G, CALIGIANA G. A CAD-augmented reality integrated environment for assembly sequence check and interactive validation[J]. Concurrent Engineering, 2004, 12(1): 67-77.
- [25] 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.
- [26] HUANG G S, TUNG C K, LIN H C, et al. Inverse kinematics analysis trajectory planning for a robot arm[C]//Proceedings of the 2011 8th Asian Control Conference (ASCC). Kaohsiung: IEEE, 2011: 965-970.

作者简介



金紫凤 女, 1995年生, 浙江嘉善人. 现为复旦大学计算机科学技术学院硕士研究生. 主要研究方向为智能机器人.
E-mail: zfjin18@fudan.edu.cn



潘思聪 男, 1996年生, 上海人. 现为复旦大学计算机科学技术学院硕士研究生. 主要研究方向为智能机器人.
E-mail: 18210240033@fudan.edu.cn



危辉(通讯作者) 男, 1971年生, 江西南昌人. 现为复旦大学计算机科学技术学院教授、博士生导师. 主要研究方向为人工智能理论与技术、认知科学.
E-mail: weihui@fudan.edu.cn