

• 机械工程 •

DOI:10.12454/j.jsuese.202400426



面向无序分拣场景的工件6D位姿检测方法

曹学鹏¹, 李鑫^{1,2}, 冯艳丽², 石瑞¹, 葛天焯¹, 张新荣¹, 赵睿英¹

(1. 长安大学 道路施工技术与装备教育部重点实验室, 陕西 西安 710064; 2. 西安航天时代精密机电有限公司, 陕西 西安 710100)

摘要:目标6D位姿检测是实现机器人自主抓取的关键。为克服传统点对识别(PPF)方法检测性能差、耗时及难以检测到多平面特征工件的6D位姿等不足,提出面向无序分拣场景的工件6D位姿检测方法。首先,基于模型平面点分布筛选多平面特征工件,提取其边界特征进行6D位姿检测,并在多视点下提取模型点对以去除冗余点对,提高算法识别速度。其次,匹配场景与模型间的点对特征,利用快速投票方案获取无序场景中目标的位姿假设集合。接下来,通过位姿验证筛选方法,剔除重复和误匹配位姿,实现目标多实例位姿的粗略估计,并借助迭代最近点(ICP)算法完成目标位姿的精确估计。实验结果表明:在无序仿真场景中,单次识别时间小于等于1.15 s,平均平移偏差小于等于0.95 mm,平均旋转误差小于等于1.56°;在实际场景中,平均识别成功率为95.82%,平均单次识别时间为1.11 s。综上,该6D位姿检测方法在保证识别效率的同时兼顾了位姿估计精度,并在识别精度和速度上均优于同类算法,为机器人的精准抓取的实现提供了有力的保障。

关键词:无序场景;6D位姿检测;点对特征;位姿估计精度;识别率

中图分类号:TP242.2

文献标志码:A

文章编号:2096-3246(2025)05-0298-11

6D位姿为物体在3维空间中的位置和朝向,包括3个自由度的位移和3个自由度的旋转。近年来,工业机器人6D位姿检测技术在工业生产得到了广泛的应用,然而,在目标散乱、相互堆叠的场景中,目标物体的6D位姿检测面临极大挑战。与其他目标识别定位方法相比,基于点对特征的识别算法在工业环境中展现出良好的应用前景。

Drost等^[1]提出基于点对特征的识别算法,在定向点对之间定义4维特征向量,并将其离散化来计算场景和模型点对之间的对应关系,有效地将相似的点对分组在一起。对于每个场景点,在由模型对应点和围绕点法线的旋转角度定义的2维空间上,对所有相应点对关系进行投票以生成候选位姿。最后,对候选位姿进行聚类 and 排序以获得最终估计结果。Choi等^[2]遵循该算法框架,设计了点到线、线到线等功能,提高点对识别算法检测边缘的能力。为了减少背景和传感器噪声的影响,Hinterstoisser等^[3]提出一种对噪声更具鲁棒性的两球采样和投票方法。郁梦辉等^[4]提出新的基

于曲率点对特征的3维识别算法。该算法在原始点对特征的基础上,引入曲率差特征,使点对具有更强的特征描述性,提高点云配准率。Liu等^[5]针对工业环境中工件点云关键特征的缺失,提出新的基于点对特征的描述子来估计工业零件的位姿并利用2维图像对点云中缺失的关键特征进行补偿。Guo等^[6]基于对象中心和点对特征之间的相对几何关系,提出通过投票在场景的目标中心附近生成新的点,利用聚类生成位姿候选并引入匹配点对计算变换矩阵,该方法更简单便捷,对噪声、遮挡和几何简单形状具有很强的鉴别力和鲁棒性。虽然以上点对识别方法在工业生产和机器人应用中表现出良好的检测性能,不过仍存在一些不足,包括:1)对传感器噪声、严重遮挡和背景杂波较为敏感^[7-9];2)工件本身具有许多重复特征(如大平面)时,匹配性能会降低^[10-12];3)依赖于搜索大量点对并计算变换关系,识别速度仍有待提升^[13-15]。近年来,基于深度学习的6D位姿检测方法^[16-18]在检测公共3D数据集时表现出较好的识别效率和准确度。然而,该

收稿日期:2024-06-03 修回日期:2025-07-27 网络出版日期:2025-07-31

基金项目:国家自然科学基金项目(62073092);陕西省重点研发计划项目(2021ZDLGY09-02;2024GX-YBXM-164)

作者简介:曹学鹏(1982—),男,教授。研究方向:机电控制系统及工业机器人应用。E-mail:tiepeng2001@chd.edu.cn

类方法需要耗费大量的计算资源、人力成本进行数据集的制作和训练,不适用于高效率的工业场景。为此,本文面向机器人分拣系统的抓取任务,提出了一种基于点对特征的6D位姿检测方法。

本文的主要工作包括:在离线阶段,通过计算模型点云平面点分布,筛选多平面特征工件,基于边界点对特征进行6D位姿检测,并在多视点下提取模型点对特征;在线阶段,提取场景边界点并通过自适应选点方法完成场景参考点的选取,借助位姿筛选算法实现场景中工件的多目标检测;开展性能测试研究,完成球形把手、三通管、转向臂、异形管件、异形连接件、“L”形连接件等典型工件的点云数据6D位姿检测,定量评估算法的识别精度;最后,搭建机器人实验平台进行6D位姿检测实验,验证该算法在无序场景中的识别率及识别时间。

1 基本原理

基于点对特征的6D位姿检测算法分为离线建模和在线匹配两阶段,流程图如图1所示。离线阶段主要计算模型的点对特征并通过量化函数 $Q(\cdot)$ 进行离散化,得到哈希表索引,将模型点对存储在哈希表中。在线阶段主要将场景与模型的点对特征进行匹配,通过快速投票方案来估计目标的候选位姿;而后通过位姿筛选方案筛选出合理的位姿集合。对典型平面特征的工件,点对特征提取阶段则计算场景和模型点云的边界点对特征。

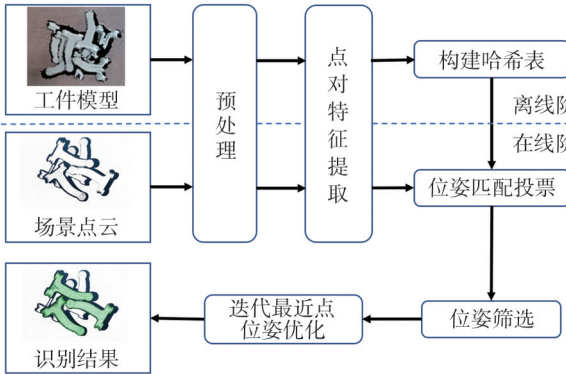


图1 6D位姿检测流程图
Fig. 1 Chart of 6D pose detection

点对特征通过两点及其法向量构建的4维特征来描述两点的相对位置和姿态。对于任意两点 m_i 和 m_r ,及其法向量 n_i 和 n_r ,定义点对特征 F 为:

$$F(m_i, m_r) = (f_1, f_2, f_3, f_4)^T \quad (1)$$

$$= (\|d\|_2, \angle(n_i, d), \angle(n_r, d), \angle(n_i, n_r))^T$$

式中: f_1 为 m_i 和 m_r 之间的欧氏距离 $\|d\|_2$, d 为点 m_i 到 m_r 的向量; f_2 为向量 d 和法向量 n_i 的夹角 $\angle(n_i, d)$; f_3 为向量 d 和法向量 n_r 的夹角 $\angle(n_r, d)$; f_4 为法向量 n_i 和 n_r 之

间的夹角 $\angle(n_i, n_r)$ 。点对特征如图2(a)所示。

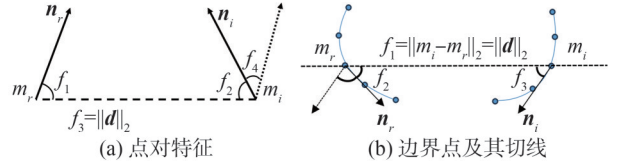


图2 点对特征

Fig. 2 Point-pair feature

对于典型平面特征的工件,边界点则更能够反映其工件形状的基本信息,为此,采用边界点对特征进行检测。通过两边界点及其切线向量构建的4维特征来描述两点的相对位置和姿态,如图2(b)所示。通过随机样本一致(RANSAC)算法^[19-21]进行边缘线段拟合,但RANSAC算法只能计算直线边界点的方向。为此,对于存在圆弧边界的工件,采用边界点方向计算方法将边界点的切线方向作为边界点对特征的方向向量。

输入点云的边界点集 P 后,边界点方向的计算方法如下:

1. for $i = 1$ to $\|P\|$ do // $\|P\|$ 表示点集 P 中点的数量
2. for $j = 1$ to $\|P\|$ do
3. if $i \neq j$ then
4. 计算经过点 p_i 和点 p_j 的直线 L_{ij} 长度;
5. 计算到直线 L_{ij} 距离小于临界距离 D_{th} 的邻近点数量 $N(p_i)$;
6. $N = N \cup (p_i, N(p_i))$ // N 为键值为 p_i 、值为 $N(p_i)$ 的集合;
7. if $\max_{1 \leq k \leq \|P\|} (N(p_k))$ then // $\max(N(p_k))$ 判断 $(N(p_k))$ 是否为 N 最大值
8. $k^* \leftarrow \arg \max_{1 \leq k \leq \|P\|} (N(p_k))$; // k^* 为到直线 L_{ij} 距离小于等于 D_{th} 的邻近点数量最多的点的索引
9. $P_o \leftarrow P_o \cup \{p_j, n(p_j)\}$; // P_o 为输出的每个边界点及对应的方向向量的集合, $n(p_j)$ 为 p_j 的方向向量
10. end
11. end
12. end

2 离线建模阶段

离线阶段进行工件类型的判别、模型点云的采样及模型点对的提取。对于模型点云 M ,利用RANSAC方法迭代计算 M 中的平面点。经多次验证,若迭代3次后平面点数量大于模型点数量的2/5,可认为工件属于多平面特征的工件,基于边界点对特征进行6D位姿检测;否则,视为非多平面特征工件,基于原始点对特

征进行检测。

2.1 体素网格均匀采样

对场景和模型点云进行降采样,但当点云轴向包围盒在 x 、 y 、 z 轴3个方向的边长相差悬殊时,传统点对识别算法中的体素网格降采样方法存在着采样点空间分布不均匀的问题^[22]。为此,对点云轴向包围盒沿 x 、 y 、 z 轴3个方向进行划分,使得每个体素近似为长度为 L_{voxel} 的正方体,然后计算每一个体素内所有点的重心,并将其作为该体素的采样值。

任一点 p 所在体素的访问索引 $(x_{\text{int}}, y_{\text{int}}, z_{\text{int}})$ 的计算式如式(2)所示。

$$\begin{cases} x_{\text{int}} = \text{int}\left(\frac{x_i - x_{\text{min}}}{L_{\text{voxel}}}\right), \\ y_{\text{int}} = \text{int}\left(\frac{y_i - y_{\text{min}}}{L_{\text{voxel}}}\right), \\ z_{\text{int}} = \text{int}\left(\frac{z_i - z_{\text{min}}}{L_{\text{voxel}}}\right) \end{cases} \quad (2)$$

式中: (x_i, y_i, z_i) 为点 p_i 的坐标值,点云轴向包围盒顶点的最小坐标值为 $(x_{\text{min}}, y_{\text{min}}, z_{\text{min}})$; $\text{int}(\cdot)$ 为向下取整函数。

2.2 模型点对特征计算

图3为模型点对的计算过程。在传统点对识别算法中,计算每两个模型点构成的点对特征并存储在哈希表中,但有些点在场景中是不可见的,如图3中的 p_a 和 p_b ,代表了标识不同的两点。当哈希表中存在这些冗余点对时,在线阶段的快速投票期间,可能将票投给错误的候选位姿,从而导致错误的位姿估计,也增加了哈希表的大小和在线匹配时间。

为解决上述点对冗余问题,分别在6个视点下提取模型的单视角点云并计算其点对特征,得到整个模型的点对描述,如图3(a)所示。对于三通管、球形把手等多平面特征工件,直接计算每个单视角点云的点对特征;而对于垫片、转向臂等非多平面特征工件则计算单视角点云的边界点对特征。为此,通过点在 x 、 y 和 z 方向的梯度和法线方向来识别每个单视角点云的边界点,提取其边界点对特征,整个模型的边界点提取结果如图3(b)所示,边界点提取过程如图3(c)所示。为记录点对及其所属的单视角模型点云,为每一个点对分配一个唯一的索引。

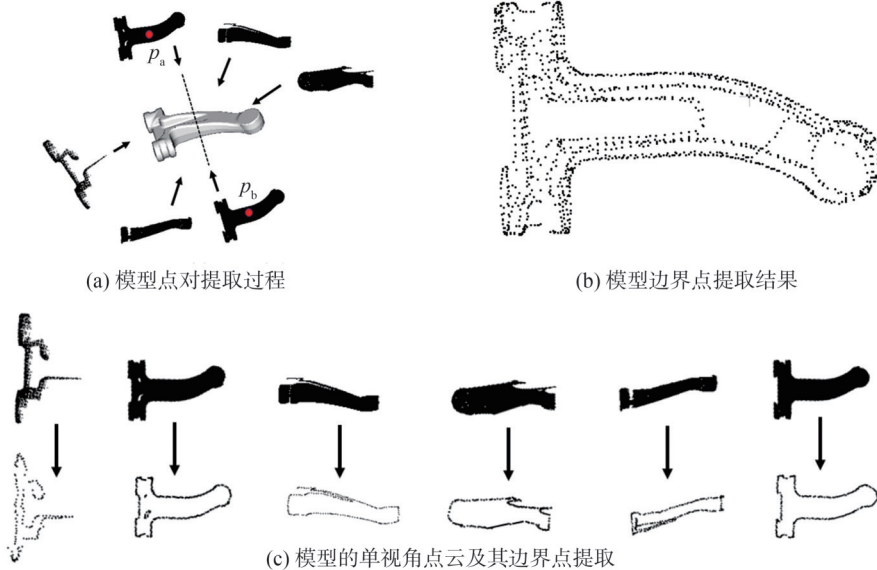


图3 模型点对的计算过程

Fig. 3 Model point-pair calculations

采用上述方法计算模型点云的点对特征时,仍可能出现点对冗余现象,重复点对示意图如图4所示。图4中, F 为点对特征向量,红色框中点对 (m_1, m_2) 和 (m_3, m_4) 实际是相同的点对,但因位于不同的单视角点云中,导致相对应的点的索引不同。为此,对于每个点对中点所在的体素索引组成的向量,当检测两个点对位于同样的位置时,只将其中一个点对保存在哈希表中。例如, m_1 和 m_2 所在的体素索引为 $(3, 4, 5)$ 和 $(6, 7, 8)$,则点对 (m_1, m_2) 所属的向量 c 为 $c = (3, 4, 5, 6, 7, 8)$ 。由于

(m_3, m_4) 与 (m_1, m_2) 位于相同的位置,故所属向量确定为 $c = (3, 4, 5, 6, 7, 8)$ 。

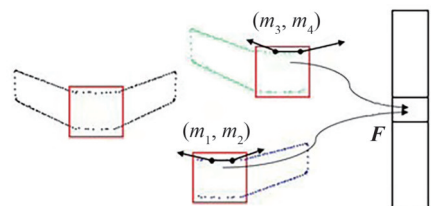


图4 重复点对示意图

Fig. 4 Repeat the point pair sketch

3 在线匹配阶段

在线阶段通过点对特征匹配投票获得目标对象的 6D 候选位姿。首先,对场景点云仍进行预处理和场景点对特征计算。考虑到工件常放置在桌子上或料箱中,为此采用分割方法从场景点云中去除这些盛放物底面对应的点云平面。而后,在哈希表中查找与场景点对相似的模型点对,在位姿参数空间内计算模型一场景点对对应的刚体变换,并在 2 维累加器中完成姿态加权投票。最后,采用位姿筛选算法剔除重复和误匹配位姿,获得目标粗估计位姿并运用 ICP 算法进行位姿优化,位姿筛选前后对比如图 5 所示。

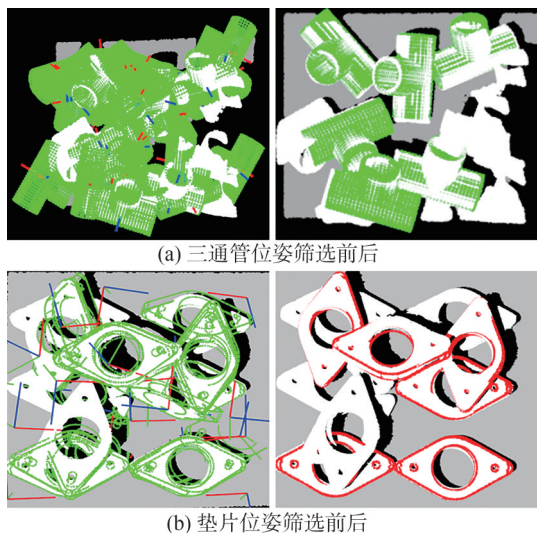


图 5 三通管和垫片的位姿筛选前后对比

Fig. 5 Three-way tube and gasket position before and after screening

3.1 场景点对特征的计算

场景点对特征计算时,场景参考点的选取对 6D 位姿检测算法的成功率及识别速度有很大影响。传统点对识别算法中根据场景点的索引,每隔 S 个点从场景中选取 1 个。不过合适的步长很难确定,需要多次试验,且因场景中不同的待识别目标形状大小不尽相同,步长值 S 也不相同。步长值 S 越小,识别速度越慢,步长值 S 过大,则会导致识别精度低。故需选取尽可能少的参考点,且确保参考点分布均匀,并满足参考点中存在待识别目标点。

为解决参考点步长 S 需多次调整的问题,采用自适应均匀选点方法,即根据场景中待识别工件的大小自适应地选取数量适当且分布均匀的参考点,确保参考点分布在待识别工件上。首先,计算目标模型点云 M 最小体积包围盒的对角线长度 d_M ,沿体素空间 x, y 两轴方向等距剔除体素,距离分别为 $\Delta x, \Delta y$ 。 $\Delta x = \Delta y = \lambda d_M$, 其中, λ 为比例系数,且 $\lambda \in [0, 1]$ 。然后,剔除不包含任何点的体素,

并在剩余的体素中找出每个 z 方向上包含点数最多的体素。最后,计算每个体素内所有点的重心,并将该体素中距离重心最近的点作为该体素的采样点。

对于三通管、球形把手等非多平面特征的工件,通过参考点方法选取场景参考点后,计算场景点对特征即可。而对于垫片、转向臂等多平面特征工件,计算场景点对之前需要提取场景点云边界点。在工业环境中,通常将点云映射到对应的 2 维图像中,利用图像边界提取算法提取图像边缘并对应到 3 维空间得到点云边界点,但经常造成折叠边缘丢失、边缘特征不完整等缺陷。针对这一问题,利用高斯映射和 K 均值聚类对点云进行聚类并计算聚类结果中各点的最近邻点的协方差矩阵的特征值,提取场景点云的边缘。图 6 为通过该方法提取边缘点的效果,包括转向臂和垫片场景点云的边界点。

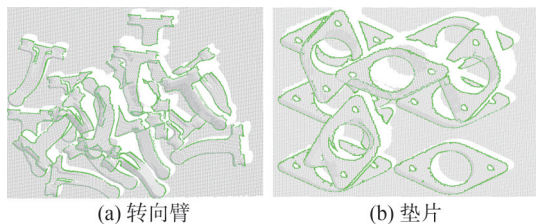


图 6 边界点提取效果

Fig. 6 Edge point extraction effect

3.2 特征匹配

特征匹配通过式(3)离散化点对特征生成量化索引,将所有相似的点对特征在哈希表中分组来解决匹配问题。对于场景点对 (s_r, s_i) ,通过在哈希表中访问其量化索引指向的位置来检索相似的模型点对 (m_r, m_i) 。然后,通过计算场景点对特征和对应的模型点对特征之间的刚体变换生成位姿假设。

$$Q(F) = \left[\frac{f_1}{I'}, \frac{f_2}{A'}, \frac{f_3}{A'}, \frac{f_4}{A'} \right]^T \quad (3)$$

式中, I' 为距离阈值, A' 为角度阈值。

刚体变换计算的具体流程为:定向点 s_r 和 m_r 在世界坐标系 X 轴对齐,参考点 s_i 和 m_i 通过绕 X 轴旋转角度 α 来对齐模型一场景点对,变换流程图如图 7 所示。

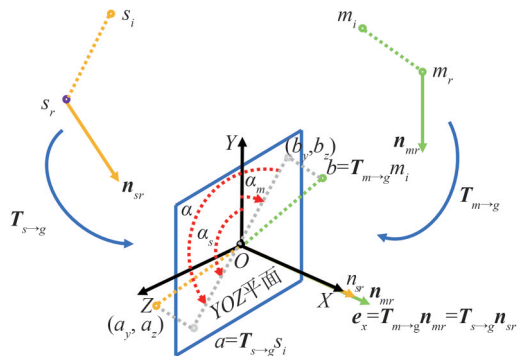


图 7 模型-场景点对变换流程

Fig. 7 Model-scene point-to-point transformation process

图 7 中, $T_{m \rightarrow g}$ 为将模型点对变换到世界坐标系的变换矩阵, $T_{s \rightarrow g}^{-1}$ 为将场景点对变换到世界坐标系的变换矩阵的逆矩阵, n_{s_v} 、 n_{m_r} 分别为点 s_r 、 m_r 的法向量, e_x 为 X 轴的方向向量, b 为 m_i 经变换矩阵 $T_{m \rightarrow g}$ 变换到世界坐标系中的点, a_z 、 b_z 分别为 a 、 b 与 Z 轴的夹角, a_y 、 b_y 分别为 a 、 b 与 Y 轴的夹角。

模型点对 (m_r, m_i) 到场景点对 (s_r, s_i) 的变换如下:

$$s_i = T_{s \rightarrow g}^{-1} R_X(\alpha) T_{m \rightarrow g} m_i \quad (4)$$

式中: $R_X(\alpha)$ 表示绕 X 轴旋转 α 的变换矩阵。为提高计算效率, 利用 YOZ 平面将旋转角 α 分成 α_m ($\alpha_m = \arctan(b_z/b_y)$) 和 α_s ($\alpha_s = \arctan(a_z/a_y)$) 两部分, 其中, $a = T_{s \rightarrow g} s_i$, $b = T_{m \rightarrow g} m_i$ 。离线建模阶段预先计算出 α_m , 将其与参考点 m_r 一起保存在哈希表中。在线匹配阶段, 对场景点对 (s_r, s_i) 计算 α_s , 得到旋转角 $\alpha = \alpha_s - \alpha_m$ 。

在点对匹配过程中, 点对距离大于模型最大尺寸的点对不属于同一目标。为此, 在投票阶段中加入式 (5) 的权重筛选标准以提高算法的匹配速度和识别精度:

$$\text{weight}(d) = \begin{cases} 1, 0 \leq d \leq d(N), d_z < \omega; \\ 0, \text{其他} \end{cases} \quad (5)$$

式中: $\text{weight}(d)$ 为权重函数; d 为点对距离; $d(N)$ 为该目标点云最大包围盒的对角线长度, $d(N) = \sqrt{d_x^2 + d_y^2 + d_z^2}$, d_x 、 d_y 和 d_z 分别是对象在 X 、 Y 和 Z 方向上的最大长度; ω 为确保目标不超过固定倾斜角的阈值。

3.3 点对投票方案

因单个模型-场景点对得到的位姿假设不足以代表模型和场景中模型的变换关系, 需要 1 组一致的对对应关系来支持相同的位姿假设。理论上, 越多的模型-场景点对对应一个刚体变换 $T_{s \rightarrow m}$, 模型和场景中模型对应的刚体变换越有可能为 $T_{s \rightarrow m}$ 。为此, 本节对相同刚体变换的点对进行分组, 通过局部坐标 (m_r, α) 和 2 维累加器迭代地为每个场景点生成位姿假设, 2 维累加器的行对应模型点集合 M 的数量; 列对应量化步长划分后的角度值, 如步长为 12, 1~12 列值为 $\pi/12 \sim \pi$ 。

选取场景参考点集合 $S(S = \{s_1, s_2, \dots, s_r\})$ 后, 对每个场景参考点 s_r 和其他场景点组成场景点对 (s_r, s_i) 进行特征匹配, 计算对应的局部坐标并在累加器中进行投票, 投票完成后累加器峰值即 s_i 的最优局部坐标 (m'_r, α') 。然后, 根据 (m'_r, α') 和式 (4) 得到候选位姿变换 (R_r, t_i) 。在所有场景参考点对参与投票后, 得到候选位姿变换集合 C , 投票过程如图 8 所示。图 8 中, $F(s_r, s_i)$ 为 s_i 和 s_r 的点对特征向量。

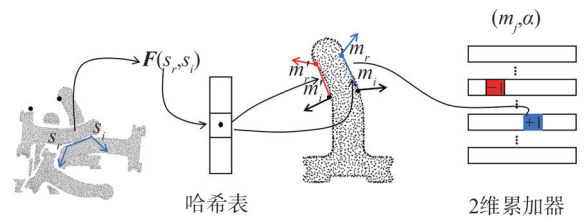


图 8 投票过程

Fig. 8 Voting process

由于传感器噪声和背景点的影响, 候选位姿集合 C 中存在一些误匹配位姿, 需要对候选位姿集合 C 进行筛选。

3.4 位姿筛选

通过候选位姿将模型点变换到场景点云的体素空间中, 建立模型点与场景点的匹配关系, 通过位姿筛选算法去除大量的误匹配位姿和重复位姿, 并筛选出场景中待识别目标的位姿粗估计结果。

首先, 建立场景点云体素空间, 并通过式 (1) 访问场景点 s_i 所在的体素 $\text{Voxel}(s_i)$ 。然后, 将所有体素的值初始化为 -1 来表示该体素内没有场景点, 并将每个场景点 s_i 所在的体素和相邻体素的值变为场景点的索引 i 。最后, 将模型点经候选位姿变换到场景空间中, 建立模型点与场景点的对应关系。体素验证过程如图 9 所示。

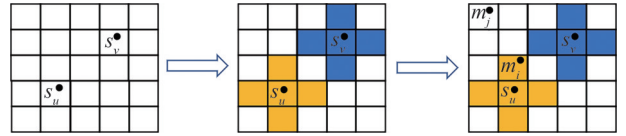


图 9 体素验证过程

Fig. 9 Voxel verification process

图 9 中, s_u 、 s_r 为模型点, 对于每个候选位姿 p_k 变换后的模型点 m_i , 计算 m_i 所在的体素值 $\text{Voxel}(m_i)$, 若 $\text{Voxel}(m_i) \neq -1$, 则认为 m_i 附近存在场景点, 否则没有与模型点 m_i 相匹配的场景点。

对每个候选位姿 p_k , 建立其对应的场景点集合 $M(p_k)$ 来存储每个模型点对应的场景点。为能够对场景中的多个目标实例进行检测, 对每个场景点 s_i 初始化一个二进制值 $\text{Exist}(s_i) = 1$, p_k 的得分计算函数 $\text{Score}(p_k)$ 计算如下:

$$\text{Score}(p_k) = \sum_{n=1}^{\text{size}(M(p_k))} \text{Exist}(M(p_k)_n) \quad (6)$$

式中, $\text{size}(M(p_k))$ 计算集合 $M(p_k)$ 的元素数量。

在计算所有候选位姿得分后, 对于单目标检测的要求而言, 一般选取得分最高的作为第 1 个粗估计位姿 p_0 。但通常在无序场景中需检测多个目标, 还需通过位姿筛选算法筛选出剩余的位姿粗估计结果。

为筛选出正确的粗估计位姿, 利用每个场景点的

二进制值 $\text{Exist}(s_i)$ 。在获取下一个粗估计位姿之前,基于每个候选位姿 p_k 的分数 $\text{Score_Old}(p_k)$ 的大小对候选位姿集合 P 进行排序。

首先,搜索属于 p_0 的场景点,并将它们的二进制值 Exist 更改为 0。然后,对每个要验证的候选位姿 p_k ,检查其所存储的场景点对应的二进制值并计算其新得分 $\text{Score_New}(p_k)$, $\text{Score_New}(p_k)$ 为二进制值仍为 1 的场景点的数量。由于更新了属于 p_0 的场景点的二进制值,在对候选位姿 $p_1, p_2, \dots, p_j, \dots, p_f$ 进行验证后,若 p_j 得分最高且 $\text{Score_New}(p_j) \geq \text{Score_New}(p_{j+1})$,则选择 $p_j (i \geq j \geq 1)$ 作为下一个粗估计位姿。输入候选位姿集合 P 和第 1 个粗匹配位姿 p_0 , 得出新的粗匹配位姿 p_j , 具体算法如下:

1. for $k=1$ to N_m do // N_m 为模型点总数
2. $m_k^p = p_0 m_k$; // 搜索属于 p_0 的场景点, m_k^p 为 m_k 经位姿 p_0 变换到场景体空间的点
3. if $\text{Voxel}(m_k^p) \neq -1$ then
4. $\text{Exist}(\text{Voxel}(m_k^p)) = 0$; // 将场景点的二进制值改为 0
5. end
6. end
7. $S_{\text{MaxNew}} = 0$; // S_{MaxNew} 为最新的最大值
8. for $k=1$ to $\text{Size}(P)$ do // $\text{Size}(\cdot)$ 返回集合元素数量
9. $\text{Score_New}(p_k) = \sum_{n=1}^{\text{size}(M(p_k))} \text{Exist}(M(p_k)_n)$; // 检查每个要验证的候选位姿 p_k 所存储的场景点对应的二进制值并计算其新得分 $\text{Score_New}(p_k)$
10. if $\text{Score_New}(p_j) > S_{\text{MaxNew}}$ then // 对候选位姿 p_1, p_2, \dots, p_j 进行验证
11. $S_{\text{MaxNew}} = \text{Score_New}(p_j)$;
12. $P_{\text{Best}} = p_j$; // P_{Best} 为最佳位姿
13. end
14. if $\text{Max_New_Score} \geq \text{Score_New}(p_{j+1})$ then
15. $p_j = \text{Best_Pose}$; // 若 p_j 得分最高且 $\text{Score_New}(p_j) \geq \text{Score_New}(p_{j+1})$, 则选择 $p_j (i \geq j \geq 1)$ 作为下一个粗估计位姿
16. break;
17. end
18. end

位姿筛选后,将候选位姿作为 ICP 算法^[23]的初始值进行优化缩小位姿估计误差。该方法不需要对模型点进行变换,也不需要重新搜索场景点,缩短了多目标实例检测的时间。

4 实验结果及分析

4.1 仿真场景检测实验

由于实际场景中难以对 6D 位姿检测方法检测精度进行定量验证。为此,对 6D 位姿检测算法进行仿真实验。计算机配置如下:CPU 型号为 Intel[®]Core[™]i7-13650HX 2.60 GHz,内存大小为 16GB RAM,操作系统为 64 位 Windows 11, GPU 为 NVIDIA GeForce RTX 4060。实验中除选取球形把手、三通管、垫片和转向臂等常见的工件外,还选取了异形管件、异形连接件和“L”形连接件为检测对象。

Blender 软件中可直接获取零件的真实位姿,方便计算本章方法的实际误差。因此,仿真实验中的工件点云通过 Blender 软件直接生成。首先,在虚拟空间中导入桌子模型;然后,让工件模型以刚体形式从桌子上方位置随机落下,并记录所有落在桌面的工件的位姿;最后,在桌面上方设置虚拟深度相机,从而得到该工件的无序堆叠场景点云。为了保证实际场景的可应用性,将部分实验工件密集堆叠放置,如图 10 所示。

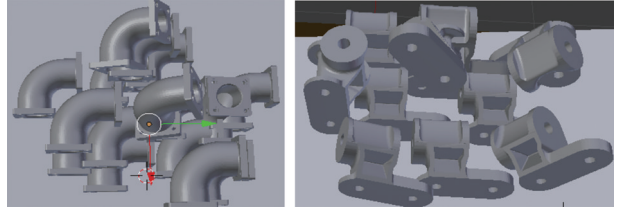


图 10 工件密集堆叠场景

Fig. 10 Scene of dense stacking of workpieces

采用上述方法对每种实验工件各生成 10 种场景点云,通过算法检测目标的 6D 位姿,结果如图 11 所示。然后,统计检测出的位姿与 Blender 场景中工件绝对位姿的平均平移误差和平均旋转误差,以及该算法的识别率和识别时间并与同类算法 PPF-Fast 算法对比,结果见表 1。其中,识别率为成功检测的工件数量占场景中工件总数的百分比。

由表 1 可知,仿真实验中 7 种工件的识别率均在 95% 以上,场景点云单次检测时间均在 1.3 s 以内。参考近 5 年同类工作,以平均平移偏差和平均旋转误差作为精度评价指标^[24-26]。仿真实验的平均平移偏差在 0.95 mm 内,平均旋转误差不超过 1.56°,故本算法在保证检测效率的同时,兼顾了位姿估计精度。并且与同类算法快速 PPF 法 (PPF-Fast) 相比,识别率与识别速度均呈现出较大优势。仿真结果表明,论文提出的 6D 位姿检测算法可有效应对无序场景中目标的 6D 位姿估计问题,为机器人抓取系统的抓取性能提供了保证。

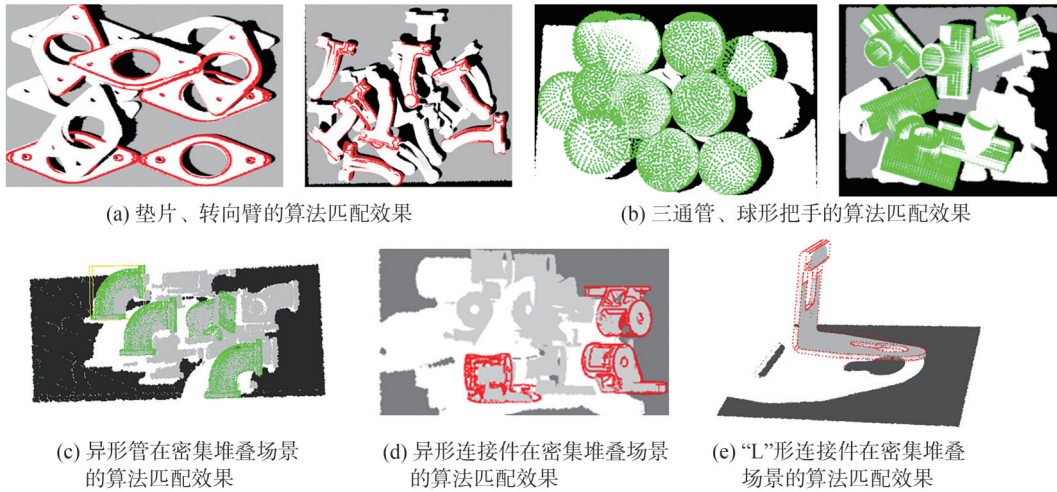


图 11 仿真场景检测结果

Fig. 11 Simulation scene detection results

表 1 6D 位姿检测结果数据统计

Tab. 1 6d position and pose detection result data statistics

零件	本文算法平均平移误差/mm	本文算法平均旋转误差/(°)	本文算法识别率/%	PPF-Fast算法识别率/%	本文算法识别时间/s	PPF-Fast算法识别时间/s
垫片	0.73	1.34	97.6	86.2	1.1	2.1
球形把手	0.84	1.29	95.2	87.3	1.2	2.6
三通管	0.75	1.19	95.8	89.5	0.9	2.8
转向臂	0.93	1.56	95.4	82.6	0.8	2.6
异形管	0.91	1.25	95.1	86.5	1.1	1.8
异形连接件	0.86	1.36	95.6	74.6	1.3	1.9
“L”形连接件	0.56	2.13	97.8	88.4	0.7	0.9

4.2 实际场景检测实验

为检验 6D 位姿检测算法在实际场景中的识别性能,通过相机对垫片、转向臂、三通管、球形把手等工件各采集 6 种不同的无序场景点云,并进行 6D 位姿检测实验。在无序场景中每种工件的个数为 8 个。每一次识别之后,采用人工的方式将识别的工件移除,以便进行下一次检测。

考虑到实际场景中难以获取每个工件质心的绝对位姿,导致无法衡量该算法的识别精度。为此,对该算法的识别率和单次场景的识别时间的平均值进行统计,来衡量算法的实际性能。不同工件在不同场景中的检测过程和算法匹配结果如图 12 所示。实验结果如表 2 所示。

由图 12(a)、(b)可知,对于垫片、转向臂,算法易受边界点提取效果的影响。相对于散乱场景,在堆叠场景中,虽然识别数量降低,但整体识别效果良好。由图 12(c)、(d)可知,对于三通管、球形把手,算法在堆叠和散乱场景中都表现出了很好的识别性能。

以上实验都是在光照合适的条件下进行。但在强光下,场景中部分工件点云不完整。为测试在强光环

境下该算法的识别性能,进行了多组实验,6D 位姿检测算法对堆叠的转向臂工件点云的识别过程如图 13 所示。

由图 13 可知,在强光条件下,工件点云的缺失导致单次识别的工件数量降低,识别次数增加,识别总时间变长,但算法的识别性能良好。

位姿检测的数据统计结果如表 2 所示。由表 2 可知,对于垫片和转向臂,在实际场景中的识别成功率分别为 95.8% 和 93.0%,相较于仿真场景略有下降,这主要是因为工件堆叠时,边界点提取效果较差,进而导致识别成功率下降;对于三通管和球形把手,实际场景中的平均识别成功率分别为 97.9% 和 95.8%,相较于仿真场景略有下降,这主要是因为工件反光时,采集到的场景点云数据存在缺失,导致其识别成功率下降。算法在实际场景中的平均识别时间为 1.19 s,相较于仿真场景相差不大。

可见,本文提出的 6D 检测方法能够有效应对无序场景中的目标检测任务,识别速度快,在强光下也能保持较好的识别性能,能够有效地应用到实际的分拣场景中。



图 12 不同场景下工件的检测过程及效果

Fig. 12 Process of workpiece detection under different scenarios

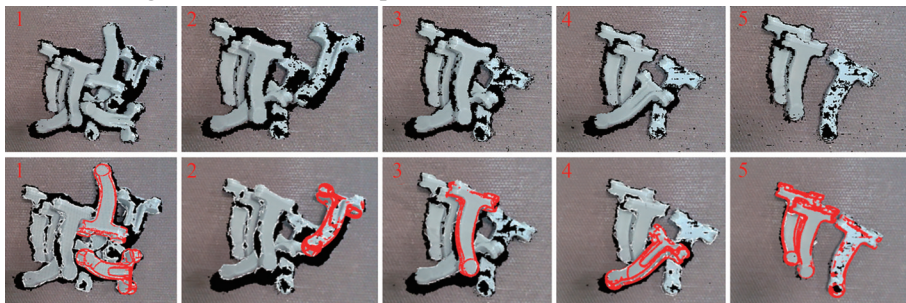


图 13 强光条件下转向臂的检测过程和匹配效果

Fig. 13 Detection process and matching effect of steering arm under strong light

表 2 位姿检测实验的数据统计

Tab. 2 Data statistics of the pose detection experiment

零件	识别成功率/%	单次场景识别时间/s
垫片	95.80	1.10
球形把手	97.90	1.07
三通管	95.80	1.14
转向臂	93.75	1.13

5 结 论

本文研究了无序场景中工件的 6D 位姿检测方法,针对无序场景中的 6D 位姿检测任务提出一种改进的基于点对特征的检测方法,并通过实验进行了验证,主要结论包括:

1) 论文面向无序分拣场景提出了改进的基于点对特征的 6D 位姿检测方法。设计了多平面特征工件模型和其场景点云的边界点对特征计算方法;在多视点下提取模型点对减少点对冗余及位姿筛选算法,以

提高该方法的识别时间、识别精度及目标的识别数量等性能。

2) 搭建了无序场景点云的仿真采集环境,对该方法的识别精度、识别率、识别时间进行定量评估。结果表明:仿真场景中该算法单次识别时间均在 1.15 s 以内,平均平移误差在 0.95 mm 内,平均旋转误差不超过 1.56°,可见,该算法适用于三通管、球形把手等非多平面特征工件和垫片和转向臂等多平面特征工件,且识别精度高、速度快。

3) 完成了实际无序分拣场景的 6D 位姿检测实验。结果表明:实际场景中该算法平均识别成功率为 95.82%,单次场景平均识别时间为 1.11 s,故该 6D 位姿检测方法检测误差小、算法识别速度快,有效提高了抓取系统的成功率和效率。

参考文献:

[1] Drost B, Ulrich M, Navab N, et al. Model globally, match locally: Efficient and robust 3D object recognition[C]//Pro-

- ceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco:IEEE,2010:998–1005.
- [2] Choi C, Taguchi Y, Tuzel O, et al. Voting-based pose estimation for robotic assembly using a 3D sensor[C]//Proceedings of the 2012 IEEE International Conference on Robotics and Automation. Saint Paul:IEEE,2012:1724–1731.
- [3] Hinterstoisser S, Lepetit V, Rajkumar N, et al. Going further with point pair features[M]//Leibe B, Matas J, Sebe N, et al, eds. Lecture Notes in Computer Science. Cham: Springer International Publishing,2016:834–848.
- [4] Yu Menghui, Cui Xining, Wu Linqigao, et al. 3D recognition algorithm based on curvature point pair features[J]. Laser & Optoelectronics Progress, 2023, 60(12):222–229. [郁梦辉, 崔西宁, 吴霖琪高, 等. 基于曲率点对特征的三维识别算法[J]. 激光与光电子学进展, 2023, 60(12):222–229.]
- [5] Liu Diyi, Arai S, Miao Jiaqi, et al. Point pair feature-based pose estimation with multiple edge appearance models (PPF-MEAM) for robotic Bin picking[J]. Sensors, 2018, 18(8):2719.
- [6] Guo Jianwei, Xing Xuejun, Quan Weize, et al. Efficient center voting for object detection and 6D pose Estimation in 3D Ppoint cloud[J]. IEEE Transactions on Image Processing, 2021, 30:5072–5084.
- [7] Labbé Y, Carpentier J, Aubry M, et al. CosyPose: Consistent multi-view multi-object 6D pose estimation[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020:574–591.
- [8] Wong J M, Kee V, Le T, et al. SegICP: Integrated deep semantic segmentation and pose estimation[C]//Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Vancouver: IEEE, 2017: 5784–5789.
- [9] Birdal T, Ilic S. Point pair features based object detection and pose estimation revisited[C]//Proceedings of the 2015 International Conference on 3D Vision. Lyon: IEEE, 2015: 527–535.
- [10] Zhao Yinshuai, Wu Qingxiao, Fu Shuangfei, et al. Efficient planar object recognition and localization system based on boundary point pair feature[J]. Application Research of Computers, 2019, 36(2):601–605. [赵银帅, 吴清潇, 付双飞, 等. 基于边缘点对特征的板型物体识别与定位系统[J]. 计算机应用研究, 2019, 36(2):601–605.]
- [11] Wu Chenghei, Jiang S Y, Song Kaitai. CAD-based pose estimation for random Bin-picking of multiple objects using a RGB-D camera[C]//Proceedings of the 2015 15th International Conference on Control, Automation and Systems. Busan:IEEE,2015:1645–1649.
- [12] Yan Wu, Xu Zhihao, Zhou Xuefeng, et al. Fast object pose estimation using adaptive threshold for Bin-picking[J]. IEEE Access, 2020, 8:63055–63064.
- [13] Lu Jun, Wei Panyi, Wang Wei. 3D target recognition algorithm based on point-pair features of key points[J]. Transactions of Beijing Institute of Technology, 2022, 42(2):200–207. [陆军, 韦攀毅, 王伟. 基于关键点的点对特征三维目标识别算法[J]. 北京理工大学学报, 2022, 42(2):200–207.]
- [14] Birdal T, Ilic S. Point pair features based object detection and pose estimation revisited[C]//Proceedings of the 2015 International Conference on 3D Vision. Lyon: IEEE, 2015: 527–535.
- [15] Park K, Patten T, Vincze M. Pix2Pose: Pixel-wise coordinate regression of objects for 6D pose estimation[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019:7667–7676.
- [16] Liang Hongzhuo, Ma Xiaojian, Li Shuang, et al. PointNet-GPD: Detecting grasp configurations from point sets[C]//Proceedings of the 2019 International Conference on Robotics and Automation. Montreal: IEEE, 2019:3629–3635.
- [17] Lan Yuqing, Duan Yao, Liu Chenyi, et al. ARM3D: Attention-based relation module for indoor 3D object detection[J]. Computational Visual Media, 2022, 8(3):395–414.
- [18] Zeng Long, Lv Wei jie, Dong Zhi kai, et al. PPR-net: Accurate 6-D pose estimation in stacked scenarios[J]. IEEE Transactions on Automation Science and Engineering, 2022, 19(4):3139–3151.
- [19] Fischler M A, Bolles R C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography[M]//Readings in Computer Vision. Amsterdam: Elsevier, 1987:726–740.
- [20] Donadi I, Pretto A. KVN: Keypoints voting network with differentiable RANSAC for stereo pose estimation[EB/OL]//(2023-07-21)[2024-06-01]. 11543. <https://arxiv.org/abs/2307.11543v3>
- [21] Jiang Jie, Ling Sirui. Parallel Voting RANSAC and its implementation on FPGA[J]. Journal of Electronics & Information Technology, 2013, 39(2), 96–100. [江洁, 凌思睿. 一种投票式并行 RANSAC 算法及其 FPGA 实现[J]. 电子技术应用, 2013, 39(2), 96–100]

- [22] Liu Keping, Gao Runze, Li Yan, et al. Fast and accurate edge extraction algorithm of stacked workpiece point cloud[C]//Proceedings of the 2021 IEEE International Conference on Real-time Computing and Robotics. Xi'ning:IEEE, 2021:75–80.
- [23] Xiao Zhengtao, Gao Jian, Wu Dongqing, et al. A uniform downsampling method for three-dimensional point clouds based on voxel grids[J]. *Machinery Design & Manufacture*, 2023(8):180–184.[肖正涛,高健,吴东庆,等.一种基于体素网格的三维点云均匀降采样方法[J]. *机械设计与制造*,2023(8):180–184.]
- [24] He Yisheng, Wang Yao, Fan Haoqiang, et al. FS6D: Few-shot 6D pose estimation of novel objects[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans:IEEE,2022:6804–6814.
- [25] Peng Sida, Liu Yuan, Huang Qixing, et al. PVNet: Pixel-wise voting network for 6DoF pose estimation[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach:IEEE,2019:4556–4565.
- [26] Wang Chen, Xu Danfei, Zhu Yuke, 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, 2019: 3338–3347.

A Method of 6D Pose Detection for Workpieces in Random Sorting Scene

CAO Xuepeng¹, LI Xin^{1,2}, FENG Yanli², SHI Rui¹, GE Tianye¹, ZHANG Xinrong¹, ZHAO Ruiying¹

(1. Institute of Mechatronics, School of engineering machinery, Chang'an University, Xi'an 710064, China;

2. Xi'an Aerospace Times Precision Electromechanical Co., Ltd., Xi'an 710100, China)

Abstract:

Objective 6D pose detection is a key technology for enabling autonomous grasping in robots. Currently, traditional point-pair feature (PPF) methods face three major challenges: 1) excessive sensitivity to sensor noise, severe occlusions, and background clutter; 2) reduced matching performance when the workpiece has numerous repetitive features; 3) slow recognition speed due to the need to search many point pairs and compute transformation relationships. This study proposes a point-pair feature-based 6D pose detection method designed for robotic sorting system grasping tasks.

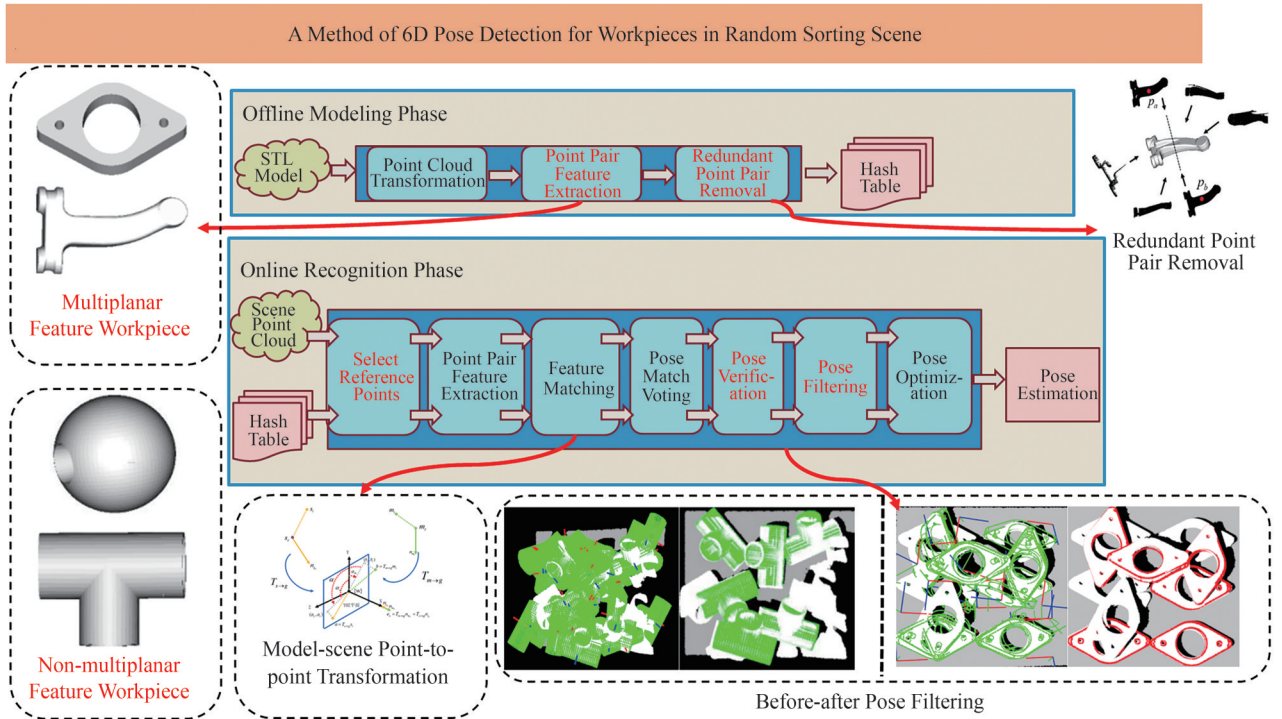
Methods Firstly, multi-plane feature workpieces were screened based on distributions of model plane points, and their boundary features were extracted for 6D pose detection. Model point pairs were extracted from multi-view points to remove redundant point pairs and improve the recognition speed of algorithms. Secondly, to further enhance recognition speed, a method was employed to extract model point pairs from multiple view-points, which helped in eliminating redundant point pairs that did not contribute to the detection process. Thirdly, the point-to-point characteristics between scenes and models were matched, and a fast voting scheme was employed to obtain pose hypothesis sets for targets in a disordered scene. Then, a pose verification and screening method was introduced to eliminate duplicate and mismatched poses, which was essential for realizing a rough estimation of multi-instance poses for the targeted workpieces. Finally, an algorithm called Iterative Closest Points (ICPs) was utilized to refine the rough estimates and achieve a more accurate estimation of the targeted poses.

Results and Discussions Experimental results showed that in the context of disordered simulation scenes, the proposed method demonstrated a single recognition time of ≤ 1.2 seconds, with an average translation deviation of ≤ 1 mm and an average rotation error of $\leq 1.56^\circ$. These results indicated a high level of precision and efficiency in pose detection. In an actual scenario, this method achieved an average recognition success rate of 95.8%, with an average single recognition time of 1.1 seconds. The high success rate and rapid speed highlighted that this method has favorable practical applicability in robotic sorting tasks. Therefore, this study highlighted that the proposed 6D pose detection method significantly outperformed the original PPF algorithm in terms of recognition speed, while also improving the accuracy of pose estimation. This advancement was crucial for the reliable and efficient operation of robotic systems in precision grasping applications. Finally, this 6D pose detection method not only ensured recognition efficiency but also accounted for the accuracy of pose estimation. This meant that recognition speed was significantly improved compared to the original PPF algorithm, providing a strong guarantee for the realization of accurate robotic grasping.

Conclusions The research presents a comprehensive approach to enhancing 6D pose detection in disordered sorting scenarios, representing a significant advancement in robotic vision and grasping technologies. The proposed method is verified as effective in both simulated environments and real-world working conditions. In addition, it demonstrates superior performance compared to existing approaches, supporting more accurate analysis in 6D pose detection applications.

Key words: disordered scene; 6D pose detection; point-to-point feature; pose estimation accuracy; recognition rate

Graphical Abstract:



(编辑 吴芝明)

引用格式: Cao Xuepeng, Li Xin, Feng Yanli, et al. A method of 6D pose detection for workpieces in random sorting scene[J]. Advanced Engineering Sciences, 2025, 57(5): 298–308. [曹学鹏, 李鑫, 冯艳丽, 等. 面向无序分拣场景的工件 6D 位姿检测方法[J]. 工程科学与技术, 2025, 57(5): 298–308.]