

DOI: 10.11830/ISSN.1000-5013.202401011



图像分割引导的散堆工件 结构光三维位姿估计

邓君^{1,2}, 李娜^{2,3}, 王亚凯^{1,2}, 高振国^{1,2}

- (1. 华侨大学 计算机科学与技术学院, 福建 厦门 361021;
2. 华侨大学 计算机视觉与机器学习重点实验室, 福建 厦门 361021;
3. 华侨大学 机电及自动化学院, 福建 厦门 361021)

摘要: 针对散堆工件场景中点云生成耗时久、位姿估计困难、多类工件混合情形难处理等问题,提出图像分割引导的散堆工件结构光三维位姿估计方法,并基于 JAKA Zu3 6-DoF 机器人开发散堆工件抓取实验系统。采用 YOLACT 模型获取散堆工件图像中的工件信息,通过自适应阈值筛选待抓取工件,利用双目结构光生成目标工件所在区域的局部点云,并基于投票匹配算法和迭代最近邻算法估计工件位姿。通过搭建的实验系统对文中方法进行测试。实验结果表明:系统完成目标工件位姿估计时间约为 3.641 s,其中,点云计算需 0.536 s,点云配准需 0.446 s;与其他方法相比,文中方法平均可缩小点云规模 44%,点云生成时间平均缩减 24%,配准成功率提升至 100%。

关键词: 双目结构光; 点云生成; 点云配准; 位姿估计; 图像分割

中图分类号: TP 391

文献标志码: A

文章编号: 1000-5013(2024)06-0696-10

Structured Light-Based 3D Pose Estimation of Piled Workpieces Guided by Image Segmentation

DENG Jun^{1,2}, LI Na^{2,3}, WANG Yakai^{1,2}, GAO Zhenguo^{1,2}

- (1. College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China;
2. Key Laboratory of Computer Vision and Machine Learning, Huaqiao University, Xiamen 361021, China;
3. College of Mechanical and Electrical Engineering, Huaqiao University, Xiamen 361021, China)

Abstract: Aiming at the problems of point cloud generation time-consuming, pose estimation difficulty, and multi class workpiece mixing difficult handling in the scene of piled workpieces, a structured light-based 3D pose estimation of piled workpieces guided by image segmentation is proposed, and an experimental system of piled workpiece picking is developed based on the JAKA Zu3 6-DoF robot. The YOLACT model is used to extract workpiece information from the piled workpieces images. The workpiece to be grabbed is filtered through adaptive threshold, the local point clouds in the area where the target workpiece is generated using binocular structured light. The workpiece pose based on the voting matching algorithm and the iterative nearest neighbor algorithm is estimated. The proposed method is tested by the constructed experimental system. The experimental results show that the system takes approximately 3.641 s to complete the target workpiece pose esti-

收稿日期: 2024-01-11

通信作者: 高振国(1976-),男,教授,博士,主要从事智能制造、机器视觉及无线自组网络等的研究。E-mail:gaohit@sina.com。

基金项目: 国家自然科学基金资助项目(62372190, 61972166); 福建省高校产学研合作资助项目(2021H6030)

mation. Among them, point cloud computing takes 0.536 s and point cloud registration takes 0.446 s. Compared with other methods, the proposed method can reduce the size of point clouds by an average of 44%, reduce the time of the point clouds generation by an average of 24%, and improve the registration success rate to 100%.

Keywords: binocular structured light; point cloud generation; point cloud registration; pose estimation; image segmentation

散堆工件分拣是工业生产中的常见任务,散堆工件分拣系统需要检测工件,对工件进行三维点云生成(三维重建)和位姿估计,再控制机械臂完成抓取。目标工件三维点云生成和位姿估计是整个系统至关重要的部分。三维点云生成方法可以分为主动式与被动式两种。被动式方法不与场景交互,通过相机采集场景图片,根据图像的纹理分布等信息恢复深度信息,进而生成点云^[1-2]。双目视觉利用左、右相机对同一场景拍摄的图像进行特征点匹配,获得视差,进而获得三维点云。该方法难点在于左、右图片像素点的匹配,匹配精确度会直接影响生成点云的效果。主动式方法通过传感器主动向场景照射信号,通过解析返回的信号计算场景的三维信息^[3-4]。主动式方法能够提供高质量的三维数据,但通常需要更多的传感器^[5]。结构光三维点云生成技术通过主动投射编码图案,更好地实现更多像素点的匹配,从而形成高密度三维点云。

对于机器人工件分拣系统,不仅需要检测工件,还需要估计工件的位姿。传统的位姿估计方法依赖于深度图或与 RGB 图像结合使用^[6-13]。这类方法难于处理散堆工件之间遮挡问题,因此大多数姿态估计方法是基于点云进行的。目前,基于点云的目标位姿估计方法主要分为基于卷积神经网络(CNN)的方法和基于特征的方法。CNN 在点云领域中已被广泛应用^[14-20],基于特征的方法十分依赖于特征点,若物体缺乏明显的特征点,使得特征提取不准确,则位姿估计会受影响^[21-33]。

上述方法在处理复杂场景时需要重建整个场景,这不仅会消耗大量时间和资源,还会生成大量冗余点,导致后续计算效率低、所需存储空间大等问题。针对这些问题,本文对图像分割引导的散堆工件结构光三维位姿估计进行研究^[34-35]。

1 系统设计

双目结构光系统中的左、右相机和投影仪分别通过 USB 接口和 HDMI 接口与计算机连接,计算机通过路由器与机器人连接。系统硬件模块,如图 1 所示。

系统架构主要由 5 个模块组成:点对特征库构建模块、YOLACT 训练模块、双目结构光局部点云生成模块、位姿估计模块、机械臂抓取模块。其中,前两个模块构成系统离线准备部分,后 3 个模块构成系统在线运行部分。系统架构,如图 2 所示。

散堆工件图像库存储了 1 000 幅利用 Labelme 标注的散堆工件图片,图片用于训练 YOLACT 网络模型。系统需要操作的各种类型工件由预先准备的系统工件类型模型决定。工件类型模型为相应类型标准工件的三维点云集合,存储于工件类型模型库中。为进行点云匹配位姿估计,计算每个工件类型的模型点云的点对特征集合,并用该集合作为相应类型工件模型的特征模型,存储于模型点对特征库中。

双目结构光局部点云生成模块包含工件筛选、结构光投影、点云计算 3 个环节。该模块通过双目系统中的左目相机拍摄散堆工件场景图像,利用 YOLACT 模型对图像进行实例分割,通过自适应阈值筛选出所有适合抓取的工件。在点云生成过程中,双目结构光系统通过投影仪将编码图案投影在散堆工件上,利用双目相机捕获带有编码图案的场景图像,生成所有适合抓取工件处的局部点云。在位姿估计模块,根据 YOLACT 输出的工件类别信息,在模型点对特征库中找到对应的模型,先后通过位姿粗配准与精配准获取工件位姿,从而指导机械臂抓取工件。工件位姿估计过程,如图 3 所示。图 3(d)中:蓝色点云为模型点云。

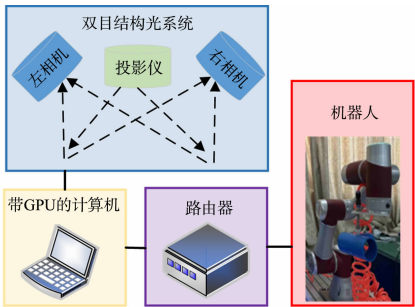


图 1 系统硬件模块
Fig. 1 System hardware modules

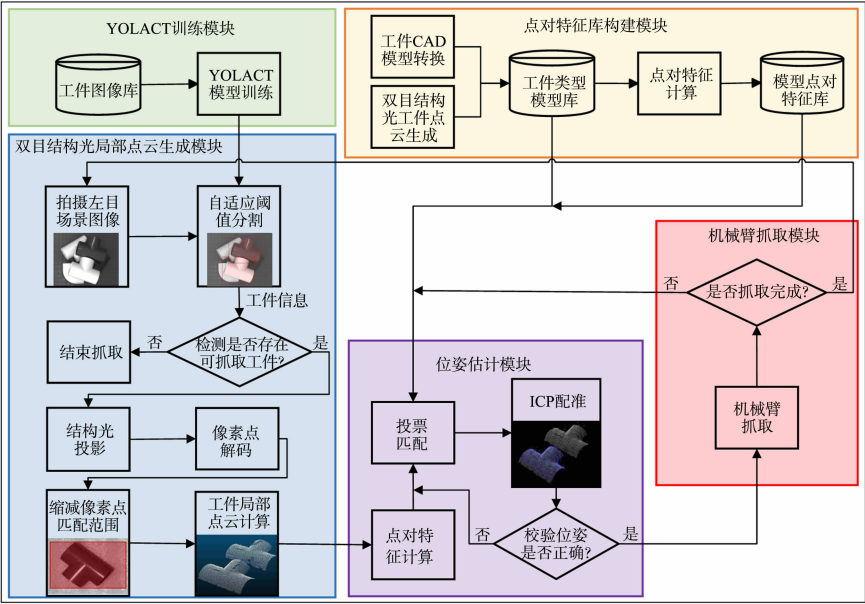
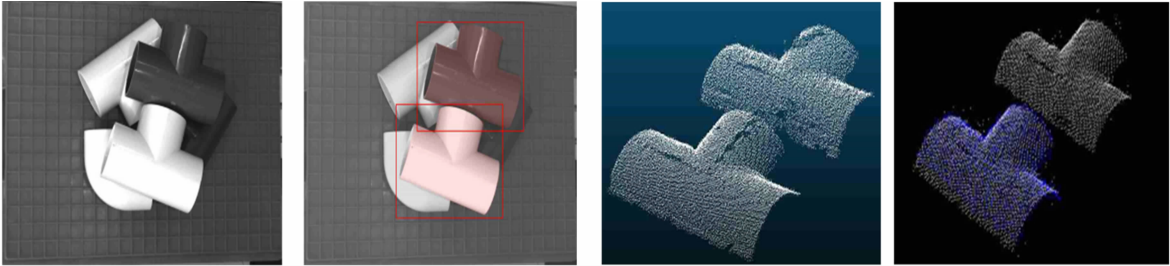


图 2 系统架构

Fig. 2 System structure



(a) 初始散堆工件场景图 (b) 工件筛选结果 (c) 生成工件局部点云 (d) 点云配准结果

图 3 工件位姿估计过程

Fig. 3 Process of workpieces pose estimation

2 双目结构光局部点云

双目结构光局部点云生成过程中,通过双目相机同步采集投影仪投射编码图案后的散堆工件场景图片。随后,通过解码计算每个像素点的绝对相位^[36],结合筛选出适合抓取工件的分割信息计算目标工件区域视差。为提高重建速度,减小点云规模,在计算目标工件区域内每行像素点对应的视差时,每隔 n 个像素点计算一次视差,生成视差图,进而转换为点云。双目结构光局部点云生成过程,如图 4 所示。

YOLACT 模型对散堆工件场景图像实例分割时,存在分割被遮挡工件的情况。为从实例分割结果中筛选出适合抓取工件的部分,引入了阈值策略。把标准工件水平放置,通过相机拍摄标准工件获取其二维图像,由 YOLACT 进行实例分割。YOLACT 把图像像素分为工件内部、工件边缘和工件外部 3 类。需要把工件边缘像素归类入工件内部和工件外部,为此计算工件内部像素点的平均灰度值 F ,再计算对应工件边缘像素点的灰度值 g 。若 $g < F/2$,则将边缘像素点归入工件外部像素点,反之,归入工件内部像素点。计算工件内部像素数量,并以此值的 70% 作为阈值来筛选适合抓取的工件。当同类型工

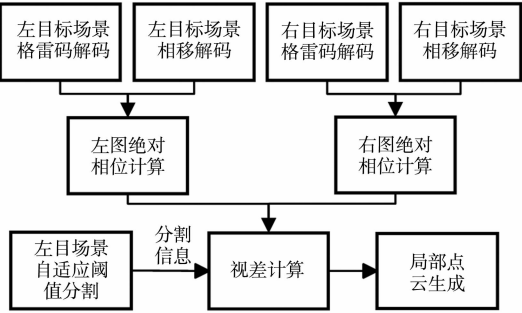


图 4 双目结构光局部点云生成过程

Fig. 4 Process of binocular structured light localpoint cloud generation

件内部像素数量超过该阈值时,认为该工件适合抓取,反之,则删除该工件的信息。散堆工件场景图像原始分割结果与筛选结果,如图 5 所示。

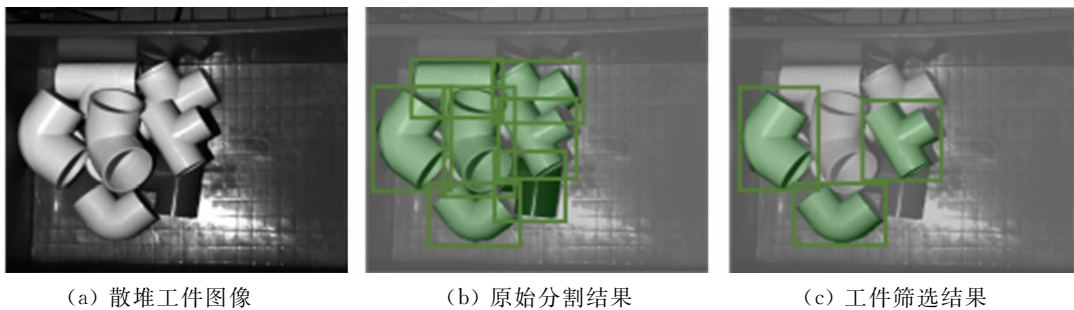


图 5 散堆工件场景图像原始分割结果与筛选结果

Fig. 5 Original segmentation results and filtered results of piled workpieces scene image

2.1 混合结构光编码方案

为减少结构光投影时间,采用格雷码与相移码相结合的结构光编码方式,共投射 p 幅格雷码图案, q 幅相移码图案,以左相机为例,左相机采集的结构光投影场景图片,如图 6 所示。

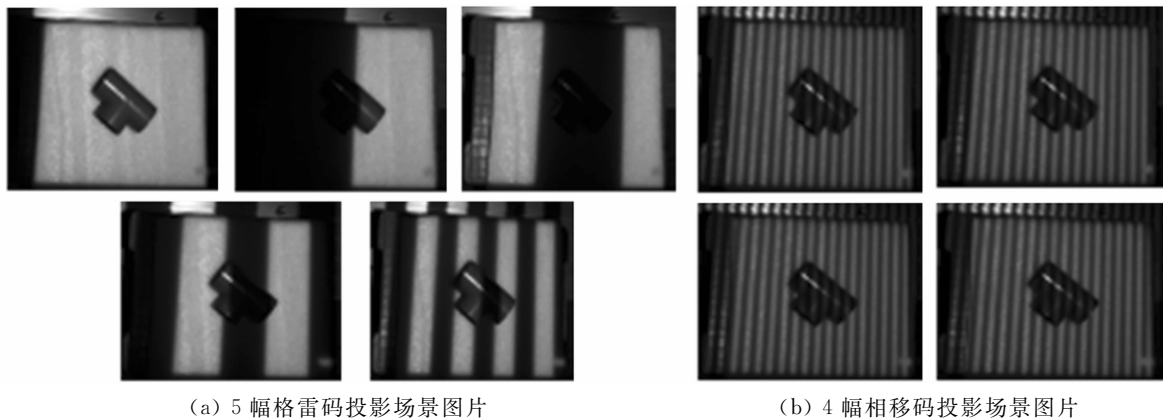


图 6 左相机采集的结构光投影场景图片

Fig. 6 Structured light projection scene images collected by left camera

经过格雷码编解码, p 位格雷码将整个图像划分为 2^p 个区域,每个区域内像素点的格雷码值相同。相移图案是周期性图案,在一个周期内,相位值是连续且唯一的。把相移码周期长度和每个区域长度设为相同,通过格雷码与相移码混合编码得到的每个像素点的绝对相位 ψ 为

$$\psi = 2k\pi + \varphi. \tag{1}$$

式(1)中: k 为像素位置的格雷码值; φ 为像素位置的相位值。

通过混合编码,可进一步区分每个区域内的像素点。格雷码解码时,为避免环境中阴影区域解码的错误二值化,通过投影正、反格雷码图案,计算带正、反格雷码图案灰度差值的方式,进行图像二值化。设 $I(x, y)$ 为投影格雷码时图像中位置 (x, y) 处的像素灰度值(光强), $I_r(x, y)$ 为投影反格雷码图案时图像中位置 (x, y) 处的光强,若 $I(x, y) \geq I_r(x, y)$,则认为该位置对应格雷码亮条纹,二值化为 1;若 $I(x, y) < I_r(x, y)$,则认为该位置对应格雷码暗条纹,二值化为 0。对拍摄的所有带格雷码图案的场景图片进行上述解码过程,就可得到每个位置的格雷码值。

相移码解码的目的是从捕获的带相移码的场景图像计算每个像素点的相位值。投射的相移图案为灰度条纹,光强表达式为

$$I(x, y) = a(x, y) + b(x, y) \cos(\varphi(x, y) + \varphi_i). \tag{2}$$

式(2)中: $a(x, y)$ 为背景光强; $b(x, y)$ 为调制强度; $\varphi(x, y)$ 表示相位值; φ_i 表示相移值。

以 4 步相移法为例,在一个 2π 周期内,相位每次移动幅度为 $\pi/2$ 。第 i 副图像的光强为 $I_1(x, y) = a(x, y) + b(x, y) \cos(\varphi(x, y))$, $I_2(x, y) = a(x, y) + b(x, y) \cos\left(\varphi(x, y) + \frac{\pi}{2}\right)$, $I_3(x, y) = a(x, y) +$

$b(x,y)\cos(\varphi(x,y)+\pi), I_4(x,y)=a(x,y)+b(x,y)\cos\left(\varphi(x,y)+\frac{3\pi}{2}\right)$ 。每个像素点的相位为

$$\varphi(x,y)=\arctan\left(\frac{I_4(x,y)-I_2(x,y)}{I_1(x,y)-I_3(x,y)}\right)。(3)$$

2.2 缩减像素点匹配搜索范围

左、右场景图案解码后,为左图像的像素点在右图像中寻找匹配点时,可以搜索所有像素点,但这样效率较低。为节省资源开销,基于 YOLACT 输出的工件分割信息,只对工件内部像素点进行匹配,并利用极线约束,将搜索范围缩小为右图像对应的极线,再通过延伸工件包围框边界,进一步将匹配范围缩小为对应极线中的一段。

为简化极线的计算,先对左、右图片进行极线校正。极线校正把两幅原始图像转化为两幅新图像。这两幅新图像中,对应极线均在与横轴平行的同一直线上,即两幅图像中互相匹配的像素点具有相同的纵坐标。

双目视觉中空间点在左、右图像的成像位置,如图 7 所示。图 7 中: P 在极线校正后的左、右图像上的坐标分别为 $(a,b),(m,n)$;两个红色方框代表极线校正后的左、右图像平面,以图像的左边界为 x 方向的零起始边;左图像中的像素位置 (a,b) 在 x 方向上相对于右图像中的像素位置 (m,n) 更靠右,即 $a>m$,且图像极线校正后使得 $b=n$,因此,把右目图像中搜索匹配点的范围缩减为第 b 行中 $[0,a]$ 区间。右图工件包围框获取过程,如图 8 所示。

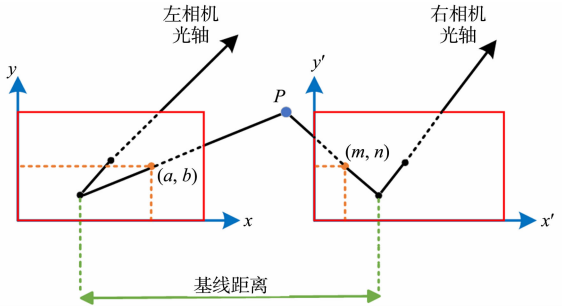
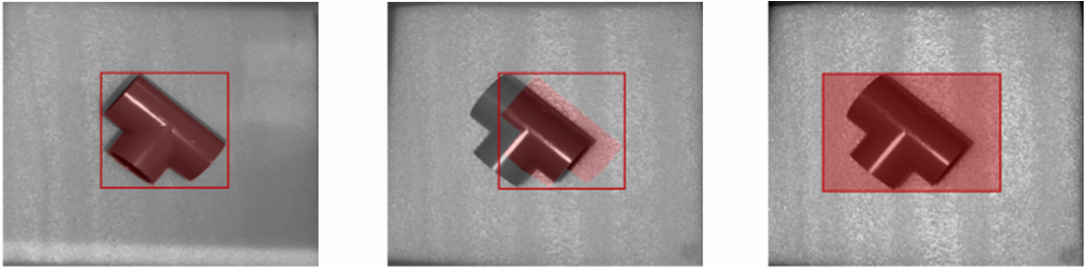


图 7 双目视觉中空间点在左、右图像的成像位置
Fig. 7 Imaging positions of spatial point in left and right images in binocular vision

由图 8 可知:左图分割得到的掩膜和包围框可以覆盖右图对应工件的大部分区域;通过计算双目系统拍摄图像的最大视差 v ,在右图上将包围框的左边界向左延伸 $(v+c)$ 个像素点,其中, c 是为保证匹配的准确性而额外延伸的像素数量;延伸后的包围框就可包含左图工件区域内像素点对应的匹配点,左图点 (a,b) 的匹配范围缩减为右图像对应极线中 $[0,a]$ 区间。通过延伸包围框边界,进一步将匹配点搜索范围缩减为 $[h,a]$ 区间,其中, h 为原包围框左边界所在像素点位置向左延伸 $(v+c)$ 个像素点后的左边界位置。



(a) 左图分割结果 (b) 分割结果用于右图 (c) 包围框延伸结果

图 8 右图工件包围框获取过程

Fig. 8 Process of obtaining bounding box of right image workpiece

3 位姿估计

3.1 模型数据库

获取工件模型点云有两种方式:一是通过工件的 CAD 模型转换为点云,将降采样后的点云作为工件模型点云;二是将双目结构光生成的工件稀疏点云通过 Cloud Compare 软件进行去噪和平滑等处理,将处理后的点云作为工件模型点云。

将工件模型点云转换为 **PPF** 特征,并用该特征作为相应类型工件模型的特征模型,以工件类型为索引构建模型点对特征库。每个模型都包含一个从 **PPF** 特征到具有相似特征的点对集合的映射,该映

射关系采用哈希表存储。

PPF 特征是一个 4 维向量, 用于表示两个空间点 p_1 和 p_2 之间的相对位置和方向关系, 即

$$PPF(p_1, p_2) = [\|d\|_2, n_1, n_2, u]。 \quad (4)$$

式(4)中: d 为 p_1 到 p_2 的方向向量; $\|d\|_2$ 为 p_1 和 p_2 之间的欧氏距离; n_1 和 n_2 分别为 p_1 和 p_2 处的法向量与向量 d 之间的夹角; u 为 p_1 和 p_2 处的两法向量之间的夹角。

3.2 基于投票匹配和迭代最近邻位姿的估计

位姿估计过程, 如图 9 所示。首先, 使用直通滤波定义点云的有效范围, 去除离群点云。接着, 通过移动最小二乘法滤波平滑点云, 以减少噪声影响, 提高匹配的稳定性。随后, 计算工件点云的点对特征集合, 并以工件类别信息为索引找到对应模型进行配准。最后, 通过投票匹配和 ICP 算法估计工件位姿。

设 M 为模型点云的点对特征, S 为环境点云的点对特征, 模型点云的点对特征($PPF_{\Theta}(p_n, p_m)$)和场景点云的点对特征($PPF_{\Omega}(q_j, q_k)$)为

$$\begin{aligned} M &= \{PPF_{\Theta}(p_1, p_2), \dots, PPF_{\Theta}(p_n, p_m)\}, \\ S &= \{PPF_{\Omega}(q_1, q_2), \dots, PPF_{\Omega}(q_j, q_k)\}。 \end{aligned} \quad (5)$$

根据 YOLACT 得到的工件种类信息, 可以在模型点对特征库中选择对应的模型, 并为每个场景点建立投票箱。比较模型点对(p_n, p_m)和场景点对(q_j, q_k)的特征向量, 若相似, 则为该模型点投票。对场景点构建的所有点对特征集合完成投票后, 将投票数(f)的最大值作为匹配点的权值 o_i , 即

$$o_i = \text{Max}(f_1, f_2, \dots, f_m)。 \quad (6)$$

为所有场景点获取匹配结果后, 可以得到多个位姿, 将相似的位姿结果分配到相同的簇中。计算簇中所有位姿投票数的总和作为该簇的总得分(W_n), 即

$$W_n = \sum_{j=1}^{n_u} \sum_{i=1}^{n_j} o_i。 \quad (7)$$

式(7)中: n_u 为簇中的位姿个数; n_j 为场景点个数。

选择总得分最高的簇, 并计算其聚类中心的对应位姿, 作为粗配准位姿结果。首先, 计算所有位姿的位置向量的平均值来求解位置均值, 其次, 计算该簇聚类中心对应的旋转矩阵(A), 即

$$A = \frac{1}{n_u} \sum_{i=1}^{n_u} w_i (u_i^T \times u_i)。 \quad (8)$$

式(8)中: u_i 为位姿的旋转矩阵; w_i 为位姿的权重。

通过投票匹配和位姿聚类, 得到了模型点云到场景点云的初始变换矩阵。使用 ICP 算法^[33]进一步精细调整, 减少点云对应点匹配的距离误差。

4 实验验证

4.1 散堆工件抓取实验系统

通过测试系统抓取工件的平均耗时、平均点云生成时间、配准成功率和平均配准时间验证方法的整体性能。平均耗时为系统抓取目标工件过程中各环节单次所耗平均时间; 平均点云生成时间与平均配准时间为获取单个工件位姿所需点云生成时间和配准时间; 配准成功率为实验中位姿估计成功次数所占比例。

搭建的散堆工件抓取实验系统(图 10)包含标准工业 6-DoF 机器人 JAKA Zu3、末端执行器、双目结构光系统、带高性能图形处理单元(GPU)的计算机。系统采用吸嘴作为末端执行器, 结合气泵、电磁阀和继电器完成抓取。吸嘴由于其工作原理, 需要接触面平滑, 且吸嘴工作时吸力较小, 因此, 将抓取点选在工件重心附近。

4.2 散堆工件抓取流程实验

工件搬移过程, 如图 11 所示。

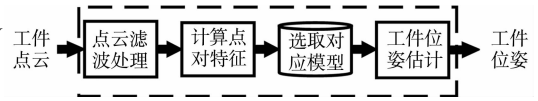


图 9 位姿估计过程

Fig. 9 Process of pose estimation

抓取全部工件仅需两次结构光点云生成(图 11 (a)):

- 1) 系统利用 YOLACT 对图片进行实例分割,并筛选出 3 个适合抓取的表层工件;
- 2) 利用双目结构光系统生成 3 个目标工件区域的局部点云,并计算工件的位姿,指导机械臂搬移工件。完成对表层 3 个工件的抓取后,系统捕获当前场景图像进行实例分割,筛选出下层的 5 个适合抓取的工件。随后,重复上述过程,完成对下层 5 个工件的搬移。搬移下层 5 个工件后,系统再次捕获场景图像进行实例

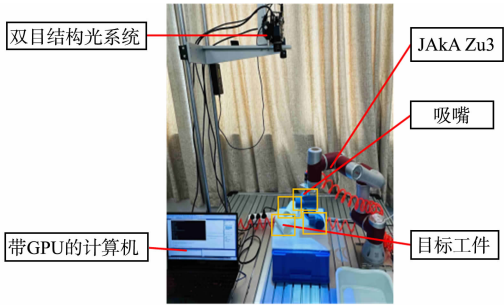
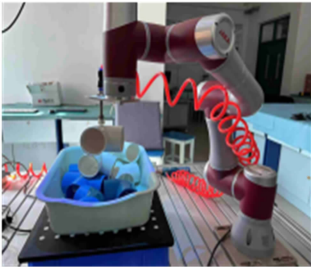


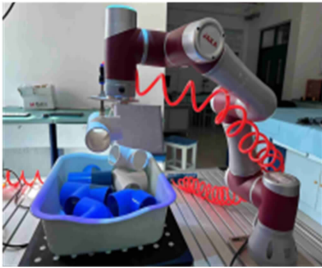
图 10 实验系统
Fig. 10 Experiment system



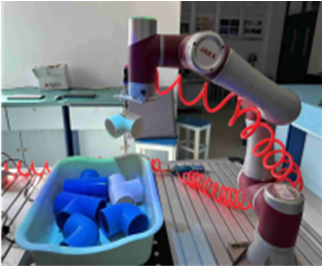
(a) 散堆工件场景图像



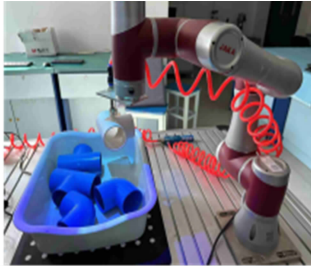
(b) 搬移上层 L 型工件



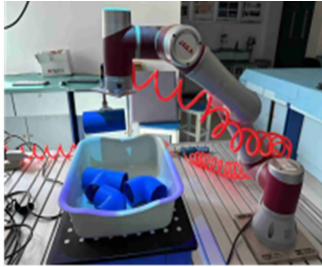
(c) 搬移上层 T 型工件



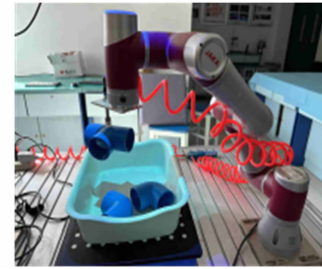
(d) 搬移上层 L 型工件



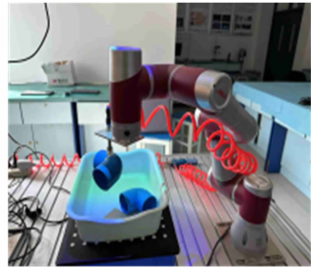
(e) 搬移下层 T 型工件 1



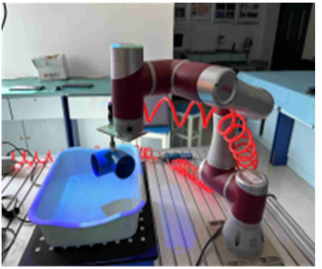
(f) 搬移下层 T 型工件 2



(g) 搬移下层 T 型工件 3



(h) 搬移下层 L 型工件 1



(i) 搬移下层 L 型工件 2

图 11 工件搬移过程

Fig. 11 Process of moving workpieces

分割,发现无适合抓取的工件,系统结束抓取任务。

点云生成环节的工件搬移过程总耗时(t)为

$$t=t_s+t_j+t_r+t_e+t_c。 \tag{9}$$

式(9)中: t_s 为工件筛选环节耗时; t_j 为结构光投影环节耗时; t_r 为计算工件点云环节耗时; t_e 为工件位姿估计环节耗时; t_c 为机械臂搬移工件环节耗时。

没有点云生成环节的工件搬移过程中,单次搬移过程总耗时 t 为机械臂搬移过程耗时。系统搬移工件各环节所耗时间,如表 1 所示。表 1 中: n 为抓取次数; $t_{rc}=t_s+t_j+t_r$; t_a 为平均时间。

机械臂搬移工件环节平均耗时为 6.024 s,远高于估计单个工件位姿平均耗时。为提升散堆工件抓取系统工作效率,令机械臂搬移工件环节与下一个工件点云的配准过程并行执行。

表 1 系统搬移工件各环节所耗时间

Tab. 1 Time consumed at each step of system in moving workpieces

n	t_s/s	t_j/s	t_r/s	t_{rc}/s	t_e/s	t_c/s	t/s
1	0.108	2.538	0.432	3.078	0.418	6.123	9.619
2	—	—	—	—	0.432	5.645	5.645
3	—	—	—	—	0.423	6.004	6.004
4	0.122	2.544	0.568	3.234	0.534	6.212	9.980
5	—	—	—	—	0.425	5.939	5.939
6	—	—	—	—	0.433	6.128	6.128
7	—	—	—	—	0.405	6.154	6.154
8	—	—	—	—	0.398	5.988	5.988
t_s	0.029	0.635	0.125	0.789	0.434	6.024	6.932

4.3 相关方法对比测试

基于 DFP^[35](DFPCP)方法、基于 PointNet⁺⁺^[15](PCP)方法及基于 DGCNN^[16](DCP)方法均利用双目结构光生成点云,使用投票匹配和迭代最近邻算法^[33]估计工件位姿。所有方法均在相同场景下实验了 50 次,4 种方法点云生成和位姿估计结果对比,如表 2 所示。表 2 中: n_r 为重建的次数; η 为配准成功率; t_{ps} 为分割时间; S 为点云规模。

表 2 4 种方法点云生成和位姿估计结果对比

Tab. 2 Comparison of point cloud generation and pose estimation results of four methods

参数	DFPCP 方法	PCP 方法	DCP 方法	文中方法
n_r	50	50	50	50
t_s/s	—	—	—	0.110
t_j/s	2.563	2.545	2.558	2.549
t_r/s	1.586	1.555	1.568	0.536
t_{ps}/s	—	0.169	0.094	—
t_{rc}/s	4.149	4.269	4.220	3.195
S	0.150	0.090	0.100	0.060
η	0.420	0.900	0.920	1.000
t_e/s	2.112	0.462	0.536	0.446
t/s	6.261	4.731	4.756	3.641

由表 2 可知:DFPCP 方法的配准成功率远低于其他方法,这是由于 DFPCP 方法生成的整个散堆工件场景点云包含大量无关点,并且 DFPCP 不能选择对应模型进行配准,而是需要依次和 L 型、T 型工件的点对特征集合进行配准,T 型工件的点对特征与 L 型工件的点对特征相似,L 型工件模型点云易配准到场景点云中 T 型工件区域,导致配准成功率低;另外 3 种方法均是针对目标工件点云进行配准,大幅度提升了配准成功率;PCP 方法和 DCP 方法需要计算完整场景点云,然后,对场景点云进行分割,以获取目标工件点云,这导致点云生成时间较长,且在散堆工件之间存在遮挡情况,两种方法都存在分割被遮挡工件的点云的情况,进而导致配准失败;文中方法结合 YOLACT 输出的目标工件的分割信息,只生成待抓取工件的点云,与其他方法相比,文中方法平均可缩小点云规模约 44%,点云生成时间平均缩短 24%,配准成功率平均提升 25%。

5 结束语

提出了一种图像分割引导的散堆工件结构光三维位姿估计方法,并基于所搭建的散堆工件抓取实验系统进行了散堆工件抓取实验。实验结果表明,该方法能快速生成所有适合抓取工件的局部点云,显著减少了点云规模和点云生成时间,双目结构光点云生成平均时间仅为 3.195 s。通过以种类为索引找到对应的模型进行基于投票匹配和 ICP 算法的位姿配准,提高了配准成功率,配准平均成功率达 100%。相较其他方法,文中方法在减小点云规模、提高点云生成速度和提高配准成功率方面表现出色,从而有效提高了抓取效率。文中方法在更复杂环境下的工件掩膜分割质量方面仍有进一步提高的空

间。对于不适合抓取的工件,系统通过机械臂对工件进行调整,确定抓取点,进而完成抓取,将是未来研究的主题。

参考文献:

- [1] SCHONBERGER J L, FRAHM J M. Structure-from-motion revisited[C]// IEEE Conference on Computer Vision and Pattern Recognition, Piscataway: IEEE Press, 2016: 4104-4113. DOI: 10.1109/CVPR. 2016. 445.
- [2] 黄及远, 李敏, 谢兵兵, 等. 双目视觉关键技术研究综述[J]. 制造业自动化, 2023, 45(5): 166-171. DOI: 1009-0134 (2023)05-0166-06.
- [3] FOIX S, ALENYA G, TORRAS C. Lock-in time-of-flight (ToF) cameras: A survey[J]. IEEE Sensors Journal, 2011, 11(9): 1917-1926. DOI: 10.1109/JSEN. 2010. 2101060.
- [4] 夏晨旭, 郝群, 张一鸣, 等. 基于结构光投影三维重建的人脸特征检测[J]. 激光与光电子学进展, 2023, 60(22): 186-192. DOI: 10.3788/LOP230620.
- [5] AHARCHI M, AITKBIR M. A review on 3D reconstruction techniques from 2D images[C]// International Conference on Smart Cities Applications, Berlin: Springer, 2020: 510-522. DOI: 10.1007/978-3-030-37629-1_37.
- [6] MAYER N, ILG E, HAUSSE R, *et al.* A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation[C]// IEEE Conference on Computer Vision and Pattern Recognition, Piscataway: IEEE Press, 2016: 4040-4048. DOI: 10.1109/CVPR. 2016. 438.
- [7] SUNDERMEYER M, MARTON Z C, DURNER M, *et al.* Implicit 3D orientation learning for 6D object detection from RGB images[C]// European Conference on Computer Vision, Cham: Springer, 2018: 699-715. DOI: 10.1007/978-3-030-01231-1_43.
- [8] 牟向伟, 孙国奇, 陈林涛, 等. 基于 Kinect 相机的多视角 RGB-D 信息融合的甜椒姿态估计研究[J]. 中国农机化学报, 2023, 44(10): 159-167. DOI: 10.13733/j.jcam. issn. 2095-5553. 2023. 10. 023.
- [9] BRACHMANN E, KRULL A, MICHEL F, *et al.* Learning 6D object pose estimation using 3D object coordinates [C]// European Conference on Computer Vision, Cham: Springer, 2014: 536-551. DOI: 10.1007/978-3-319-10605-2_35.
- [10] KRULL A, BRACHMANN E, MICHEL F, *et al.* Learning analysis-by-synthesis for 6D pose estimation in RGB-D images[C]// IEEE International Conference on Computer Vision, Piscataway: IEEE Press, 2015: 954-962. DOI: 10.1109/ICCV. 2015. 115.
- [11] 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, Piscataway: IEEE Press, 2015: 4391-4399. DOI: 10.1109/ICCV. 2015. 499.
- [12] HODAN T, HALUZA P, OBRŽÁLEK Š, *et al.* T-LESS: An RGB-D dataset for 6D pose estimation of texture-less objects[C]// IEEE Winter Conference on Applications of Computer Vision, Piscataway: IEEE Press, 2017: 880-888. DOI: 10.48550/arXiv. 1701. 05498.
- [13] JAFARI O H, MITZEL D, LEIBE B. Real-time RGB-D based people detection and tracking for mobile robots and head-worn cameras[C]// IEEE International Conference on Robotics and Automation, Piscataway: IEEE Press, 2014: 5636-5643. DOI: 10.1109/ICRA. 2014. 6907688.
- [14] QI C R, SU Hao, MO Kaichun, *et al.* PointNet: Deep learning on point sets for 3D classification and segmentation [C]// IEEE Conference on Computer Vision and Pattern Recognition, Piscataway: IEEE Press, 2017: 652-660. DOI: 10.1109/CVPR. 2017. 16.
- [15] QI C R, YI Li, SU Hao, *et al.* Pointnet⁺⁺: Deep hierarchical feature learning on point sets in a metric space[C]// Annual Conference on Neural Information Processing Systems, New York: MIT Press, 2017: 5099-5108. DOI: 10.48550/arXiv. 1706. 02413.
- [16] WANG Yue, SUN Yongbin, LIU Ziwei, *et al.* Dynamic graph CNN for learning on point clouds[J]. ACM Transactions on Graphics, 2019, 38(5): 1-12. DOI: 10.48550/arXiv. 1801. 07829.
- [17] WANG Chen, XU Danfei, ZHU Yuke, *et al.* DenseFusion: 6D object pose estimation by iterative dense fusion[C]// IEEE Conference on Computer Vision and Pattern Recognition, Piscataway: IEEE Press, 2019: 3343-3352. DOI: 10.48550/arXiv. 1901. 04780.
- [18] GAO Ge, LAURI M, WANG Yulong, *et al.* 6D object pose regression via supervised learning on point clouds[C]//

- IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2020: 3643-3649. DOI: 10.1109/ICRA40945. 2020. 9197461.
- [19] ZHUANG Chuangang, LI Shaofei, DING Han. Instance segmentation based 6D pose estimation of industrial objects using point clouds for robotic bin-picking[J]. Robotics and Computer-Integrated Manufacturing, 2023, 82: 102541. DOI: 10.1016/j.rcim. 2023. 102541.
- [20] ZHUANG Chuangang, WANG Haoyu, DING Han. Attention vote: A coarse-to-fine voting network of anchor-free 6D pose estimation on point cloud for robotic bin-picking application[J]. Robotics and Computer-Integrated Manufacturing, 2024, 86: 102671. DOI: 10.1016/J.RCIM. 2023. 102671.
- [21] PAVLAKOS G, ZHOU Xiaowei, CHAN A, *et al.* 6-DoF object pose from semantic keypoints[C]// IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2017: 2011-2018. DOI: 10.1109/ICRA. 2017. 7989233.
- [22] ROTHGANGER F, LAZEBNIK S, SCHMID C, *et al.* 3D object modeling and recognition using local affine-invariant image descriptors and multi-view spatial constraints[J]. International Journal of Computer Vision, 2006, 66(3): 231-259. DOI: 10.1007/s11263-005-3674-1.
- [23] TULSIANI S, MALIK J. Viewpoints and keypoints[C]// IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 1510-1519. DOI: 10.1109/CVPR. 2015. 7298758.
- [24] RUBLEE E, RABAU D V, KONOLIGE K, *et al.* ORB: An efficient alternative to SIFT or SURF[C]// IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2011: 2564-2571. DOI: 10.1109/ICCV. 2011. 6126544.
- [25] BAY H, TUYTELAARS T, VAN GOOL L. Surf: Speeded up robust features[C]// European Conference on Computer Vision. Berlin: Springer, 2006: 404-417. DOI: 10.1007/11744023_32.
- [26] RUSU R B, BLODOW N, BEETZ M. Fast point feature histograms (FPFH) for 3D registration[C]// IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2009: 3212-3217. DOI: 10.1109/ROBOT. 2009. 5152473.
- [27] TOMBARI F, SALT S, DI STEFANO L. Unique signatures of histograms for local surface description[C]// European Conference on Computer Vision. Berlin: Springer, 2010: 356-369. DOI: 10.1007/978-3-642-15558-1_26.
- [28] VIDAL J, LIN C Y, MARTÍ R. 6D pose estimation using an improved method based on point pair features[C]// International Conference on Control, Automation, and Robotics. Piscataway: IEEE Press, 2018: 405-409. DOI: 10.1109/ICCAR. 2018. 8384709.
- [29] SHARP G C, LEE S W, WEHE D K. ICP registration using invariant features[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(1): 90-102. DOI: 10.1109/34. 982886.
- [30] YANG Jiaolong, LI Hongdong, CAMPBELL D, *et al.* Go-ICP: A globally optimal solution to 3D ICP point-set registration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(11): 2241-2254. DOI: 10.1109/TPAMI. 2015. 2513405.
- [31] 汪霖, 郭佳琛, 张璞, 等. 基于改进 ICP 算法的三维点云刚体配准方法[J]. 西北大学学报(自然科学版), 2021, 51(2): 183-190. DOI: 10.16152/j.cnki.xdxbr. 2021-02-002.
- [32] 荆路, 武斌, 方锡禄. 基于 SIFT 特征点结合 ICP 的点云配准方法[J]. 激光与红外, 2021, 51(7): 944-950. DOI: 10.3969/j.issn.1001-5078. 2021. 07. 019.
- [33] ZHANG Jing, YIN Baoqun, XIAO Xianpeng, *et al.* 3D detection and 6D pose estimation of texture-less objects for robot grasping[C]// International Conference on Control and Robotics Engineering. Piscataway: IEEE Press, 2021: 33-38. DOI: 10.1109/ICCRE51898. 2021. 9435702.
- [34] BOLYA D, ZHOU Chong, XIAO Fanyi, *et al.* YOLACT: Real-time instance segmentation[C]// IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 9157-9166. DOI: 10.1109/ICCV. 2019. 00925.
- [35] LI Beiwen, AN Yatong, CAPPELLERI D, *et al.* High-accuracy, high-speed 3D structured light imaging techniques and potential applications to intelligent robotics[J]. International Journal of Intelligent Robotics and Applications, 2017, 1(1): 86-103. DOI: 10.1007/s41315-016-0001-7.
- [36] 熊宗刚. 结构光双目三维成像关键技术研究[D]. 成都: 电子科技大学, 2021.