

基于点云特征提取的电力刀闸 三维状态识别方法

廖 华¹, 周 庆², 蔡 宇¹, 袁卫义¹, 陈妍华³

(1. 中国南方电网有限责任公司超高压输电公司南宁监控中心, 广西 南宁 530028; 2. 华雁智能科技(集团)股份有限公司, 四川 成都 610000; 3. 华南理工大学电力学院, 广东 广州 510641)

摘 要:电力刀闸是电力系统基本组成部件, 系统运维时刀闸状态往往需要人工多次确认。在此背景下, 提出一种新的刀闸状态识别方法。首先, 将输入的三维刀闸图像转换为彩色点云数据, 并从彩色点云数据中提取场景特征; 然后, 利用现有的方向直方图颜色特征描述算法, 基于局部纹理和形状信息构建识别特征的描述符, 并根据给定提取的特征描述符, 执行两阶段匹配过程以找到场景和目标的彩色点云模型之间的对应关系; 其次, 使用霍夫投票算法过滤对应集中的匹配误差, 并估计刀闸的初始三维姿态; 最后, 姿态估计阶段采用随机样本一致性和假设验证算法来改进初始姿态, 且过滤带有错误假设的不良估计结果。实验结果表明, 该方法不仅能成功识别复杂电力场景中的刀闸部件, 而且能准确估计目标的三维姿态信息。

关 键 词:电力刀闸; 三维图像识别; 特征提取

DOI:10.19781/j.issn.1673-9140.2022.03.023 中图分类号:TM863 文章编号:1673-9140(2022)03-0190-09

Research on 3D state recognition method of power switch based on point cloud feature extraction

LIAO Hua¹, ZHOU Qing², CAI Yu¹, YUAN Weiyi¹, CHEN Yanhua³

(1. Nanning Monitoring Center of EHV Transmission Company, China Southern Power Grid Co., Ltd., Nanning 530028, China; 2. Huayan Intelligent Technology (Group) Co., Ltd., Chengdu 610000, China; 3. South China University of Technology, Guangzhou 510641, China)

Abstract: The power switch is a basic component of the power system. The switch status during system operation and maintenance often needs to be manually confirmed for multiple times. Under the background, a new method for recognizing the state of the knife switch is proposed. Firstly, the input 3D knife gate image is converted into color point cloud data, and the scene features are extracted from the color point cloud data. Then, the existing direction histogram color feature description algorithm is employed to construct the recognition feature of descriptor based on the local texture and shape information. With the given extracted feature descriptors, a two-stage matching process is performed to find the correspondence between the scene and the color point cloud model of the target. Secondly, the Hough voting algorithm is utilized to filter the matching errors in the corresponding set and estimate the initial three-

收稿日期:2021-06-09; 修回日期:2021-07-20

基金项目:成都市重大科技创新项目(2019-YU08-0067-GG)

通信作者:廖 华(1981-),男,硕士,高级工程师,主要从事电力设备变电运维研究;E-mail:liaobaby2000@qq.com

dimensional posture of the knife gate. Finally, in the pose estimation stage, the random sample consistency and hypothesis verification algorithms are used to improve the initial pose and filter out bad estimation results with incorrect assumptions. The experimental results show that the method can successfully identify the switch parts in complex power scenes, and can accurately estimate the three-dimensional posture information of the target.

Key words: power switch; 3D image recognition; feature extraction

刀闸状态确认是变电站运行人员的重要工作,设备停电维护、新设备投产等都需要进行状态确认,随着操作任务的几何级数增长和运行人员数量不足的矛盾日益增加,而基于 3D 图像识别的智能刀闸状态识别方法能有效地解决这一矛盾。3D 图像识别技术不仅在复杂环境中具有较高的目标识别率,而且可以精确地估计目标的 3D 姿态信息^[1]。根据特征描述符的构造方式,目前 3D 图像识别可分为全局方法和局部方法。全局方法基于整个目标的表面几何形状提取特征描述符,而局部方法提取和描述目标的局部区域中的特征点^[2-3]。

在全局方法的研究中,文献[4]中提出了一个综合的目标分类方法,采用基于全局的表面描述方法,当目标点云聚类被分割时,全局表面描述方法可以产生对应的全局聚类标记;文献[5]提出了一种称为视点特征直方图(viewpoint feature histog, VFH)的 3D 特征描述符,对几何和视点线索编码,用于目标识别和姿态估计。然而,VFH 描述符要求物体只能处于轻度杂乱状态,因此在复杂的现实环境中不能很好地工作。

与 3D 全局描述符相比,3D 局部描述符在杂乱场景中鲁棒性更强,因为其可以提供更多关于 3D 目标或场景的几何信息^[6-8]。文献[9]提出了点特征直方图(point feature histograms, PFH)描述符来处理点匹配问题,然而 PFH 描述方法计算量很大,并且难以实时执行;文献[10]中提出了方向梯度直方图(histogram of oriented gradient, HOG)描述符的特征提取方法,在特征点周围的球形区域内编码表面信息,通过将所有局部直方图拼接在一起,获得镜头描述符,而且镜头描述符是旋转不变的,对有噪声的场景具有较强鲁棒性。

3D 关键点匹配是局部方法中的一个重要研究内容。文献[11]提出了一种鲁棒的点匹配算法,称为向量场一致性算法(vector field consensus, VFC),该算法通过在 2 组点之间插值向量场来求解

对应关系,以估计匹配满足非参数几何约束内点的一致性;文献[12]中提出了一种基于 L2 极小化变换的估计算法,并将其应用于非刚性匹配问题。

在目标姿态估计方面,文献[13]通过将 3D 模型的投影拟合到图像中的目标轮廓来迭代计算最终的目标姿态;文献[14]提出了一种基于线性回归的姿态估计方法,该方法将高维特征向量直接映射到头部姿态角和周围偏移的联合空间。

上述研究中大部分只是针对了 3D 特征描述、3D 关键点匹配、3D 目标识别、姿态估计等单一方面的问题解决,而且未针对 3D 识别技术在电力部件识别方面的应用。因此,本文提出一种新的 3D 刀闸状态识别方法,同时处理复杂电气环境中的刀闸识别和姿态估计任务。在目标识别过程中,开始计算特征描述之前,首先,使用点云分割方法来获得可能的目标聚类;然后,执行基于关键点的两阶段匹配过程,以加速寻找当前场景的目标簇和目标的彩色点云模型之间对应关系的计算;其次,采用霍夫投票算法来过滤对应集中的匹配误差,并估计目标的初始姿态,在姿态估计过程中,利用 RANSAC 和假设验证算法对初始姿态进行优化,滤除不良估计结果;最后,实验结果验证该方法在复杂真实场景下的目标识别性能。

1 系统结构

本文所提出方法的系统结构如图 1 所示,该系统由目标识别模块和姿态估计模块组成。首先,场景的颜色和深度图像由 RGB-D 相机获取,以产生彩色点云模型。目标识别模块检测场景中的前景目标,并通过匹配前景目标和数据库中记录的目标模型之间的特征描述符来识别目标。最后,目标的类型和其在场景中的初始姿态通过霍夫投票算法同时获得。

为了提高姿态估计的精度,通过姿态估计模块进一步细化目标的姿态评估,该模块采用 RANSAC

算法来优化目标在参考系中的相对姿态,并采用一个假设验证算法来获得在预定义的姿态假设下的最佳目标姿态估计结果。

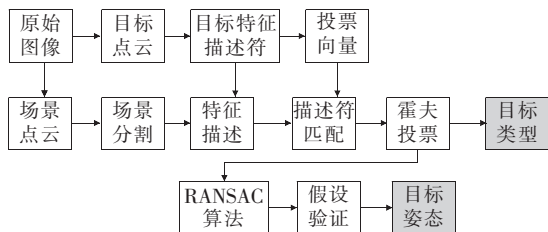


图1 所提出方法的系统结构

Figure 1 System structure of the proposed method

2 目标识别

本文所提出方法的目标识别模块由4个单元组成:场景分割、特征描述、描述符匹配和霍夫投票。当从相机捕获场景的彩色点云时,场景分割单元用于移除点云中的背景点;特征描述单元用于构造场景中所有前景目标的特征描述符;描述符匹配和霍夫投票单元识别目标及其初始姿态信息。

2.1 场景分割

场景分割是目标检测和识别的重要任务。场景分割的目的是分离场景的前景目标和平面背景区域,以提高后续特征提取和描述过程的计算效率。首先,场景的点云被分成多个平面区域,这些平面区域被认为是场景的背景区域;然后,场景的前景区域通过欧几里德距离标准进行聚类,以提取每个背景平面上的目标;最后,聚集的目标点被合并以形成前景点云,用作特征描述单元的输入。在复杂场景中,可能有多个平面需要分割。因此,所提出方法的场景分割单元采用基于连通分量的多平面分割方法来处理这种情况。

从规则2D网格采样的3D点云数据分割成有组织的点云模型。在获得连通分量的标记图像后,通过标记图像中较大的连通区域来检测场景中的平面背景。这些区域通常对应于场景中的墙平面、地平面或桌面平面,可以用平面方程 $n_x x + n_y y + n_z z + n_d = 0$ 表示,其中 $\mathbf{n} = [n_x, n_y, n_z]^T$ 是平面的表面法向量,法向量 \mathbf{n} 和平面上的点 $\mathbf{d} = [x, y, z]^T$ 之间的投影距离为

$$n_d = [\mathbf{x}, \mathbf{y}, \mathbf{z}]^T \mathbf{n} \quad (1)$$

采用了在文献[15]中提出的实时表面法向量估计算法来有效地估计场景中每个平面的法向量。最后,组织点云中的数据点 $\mathbf{p} = [\mathbf{d}^T, \mathbf{n}^T, n_d]^T$,其中投影距离 n_d 由式(1)的内积计算。

用 $\mathbf{p}_1, \mathbf{p}_2$ 表示点云的2个相邻数据点。然后使用2个距离度量 d_θ 和 d_d 来评估 $\mathbf{p}_1, \mathbf{p}_2$ 之间的相似性,即

$$d_\theta(\mathbf{p}_1, \mathbf{p}_2) = \mathbf{n}_1^T \mathbf{n}_2 \quad (2)$$

$$d_d(\mathbf{p}_1, \mathbf{p}_2) = |n_d^1 - n_d^2| \quad (3)$$

式(2)、(3)中 $\mathbf{n}_1, \mathbf{n}_2$ 分别为 $\mathbf{p}_1, \mathbf{p}_2$ 的法向量; n_d^1, n_d^2 分别为 $\mathbf{p}_1, \mathbf{p}_2$ 的垂直距离; d_θ, d_d 为 \mathbf{p}_1 和 \mathbf{p}_2 之间的差异。点云中的平面的连通分量为

$$C_p(\mathbf{p}_1, \mathbf{p}_2) = \begin{cases} \text{true,} & \text{if } d_\theta < t_\theta, d_d < t_d \\ \text{false,} & \text{otherwise} \end{cases} \quad (4)$$

式中 t_θ, t_d 分别为评估 \mathbf{p}_1 和 \mathbf{p}_2 之间的方位角和投影距离的相似性阈值。最后,通过3D连通分量标记算法^[16]求解式(4),生成背景平面的标记点云 L 。

在分离背景平面之后,对其余数据点应用欧几里德聚类过程来对场景中的目标聚类进行分类。 \mathbf{p}_1 和 \mathbf{p}_2 之间的欧几里德距离,用 $d_e(\mathbf{p}_1, \mathbf{p}_2) = \|\mathbf{p}_1 - \mathbf{p}_2\|$ 表示。 Ω_p 是在点云中检测到的平面标签集,然后根据标记的点云 L 来确定场景中目标的连接组件,使得:

$$C_{\text{obj}}(\mathbf{p}_1, \mathbf{p}_2) = \begin{cases} \text{false,} & \text{if } L(\mathbf{p}_1) \in \Omega_p, L(\mathbf{p}_2) \in \Omega_p, d_e(\mathbf{p}_1, \mathbf{p}_2) > t_e \\ \text{true,} & \text{otherwise} \end{cases} \quad (5)$$

式中 t_e 为 \mathbf{p}_1 和 \mathbf{p}_2 之间的欧几里德距离的阈值。如果2个数据点之间的欧氏距离大于 t_e ,或者其中一个点属于标签集 Ω_p ,那么这2个数据点不属于同一个目标簇。类似地,将3D连通分量标记算法应用于式(5)获取标记的目标簇。

2.2 特征提取和描述

特征提取操作仅适用于从先前场景分割过程中获得的目标簇数据点。在本文中,使用均匀采样(uniform sampling, US)方法对每个对象簇的数据点进行采样,并将剩余的点作为每个检测对象的特征点。均匀采样方法将目标的点云划分成多个立方体

区域,并将每个立方体的质心点作为目标的特征点。将立方体区域的大小设置为 $1\text{ cm} \times 1\text{ cm} \times 1\text{ cm}$ 在测试中效果较好,均匀采样方法的主要优点是可以大大地降低过程的计算成本。然后对检测目标的每个特征点执行特征描述操作。在这一步中,使用现有的 CSHOT 特征描述算法,该算法用特征点的相邻点来表征特征点,以产生如图 2 所示的组合纹理形状 3D 描述符。 p_f 表示被检测目标的特征点,通过直方图的多个特征来建立特征点 p_f 的 CSHOT 描述符,即

$$D(p_f) = \bigcup_{i=1}^m S_{f_i}^{G_i}(p_f) \quad (6)$$

式中 m 为直方图的特征数; $S_{f_i}^{G_i}(p)$ 为给定特征点 p 相对于第 i 个属性函数 G_i 和第 i 个度量函数 f_i 的直方图的特征。CSHOT 描述符包含 2 个直方图特征。第 1 个是形状相关测量的直方图特征,将其 G_1 函数和 f_1 函数分别定义为特征点的法向量和法向量的内积,使得:

$$f_1(G_1(p), G_1(q)) = \mathbf{n}_p^\top \mathbf{n}_q \quad (7)$$

式中 q 为特征点 p 的相邻点。第 2 个是纹理测量相关的特征,将 G_2 和 f_2 函数分别定义为特征点的 HSV 颜色纹理和色差的一个范数值,即

$$f_2(G_2(p), G_2(q)) = \sum_{j=1}^3 |h_j(p) - h_j(q)| \quad (8)$$

式中 $h(p)$ 为与特征点 p 相关联的 HSV 颜色三元组; j 为颜色通道的索引。

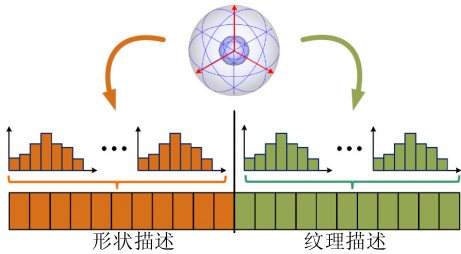


图 2 CSHOT 特征描述符

Figure 2 CSHOT feature descriptor diagram

2.3 描述符匹配

在提取目标簇中每个特征点的 CSHOT 描述符之后,执行描述符匹配过程以找到检测到的目标点云和记录的模型点云之间的三维对应关系。一般这个问题通过基于 $k-d$ 树技术的 NNS 算法来解决。然而,本文中使用的 CSHOT 描述符的维数为

1 344,由于维数灾难问题,这大大降低了基于 $k-d$ 树的神经网络系统(network nervous system, NNS)算法的匹配性能。为了解决这个问题,本文在基于 $k-d$ 树的 NNS 算法中增加了 2 种选项。第 1 种是使用多个随机化的 $k-d$ 树,而不是传统的 $k-d$ 树方法。这种设计有助于加速 NNS 过程,并且可以通过近似最近邻的快速库实现,该库提供了在高维特征描述符的大数据集中解决 NNS 问题的有效方法。

第 2 种是通过采用两阶段 CSHOT 描述符匹配算法来降低 NNS 过程的计算复杂度,该算法将 NNS 过程分为初步候选搜索阶段(preliminary candidate search, PCS)和最佳匹配确定阶段(best match determination, BMD)。在 PCS 阶段,仅采用 CSHOT 描述符的形状描述来搜索每个特征点的初步匹配候选。在 BMD 阶段,通过评估完整 CSHOT 描述符的相似性,从初步匹配候选中选择每个特征点的最佳匹配。提出的两阶段匹配算法的主要目的是通过不太复杂的匹配操作高效地搜索候选特征匹配,然后以更严格的方式从候选匹配中选出最佳匹配。因此,当特征点的数量快速增加时, NNS 过程的处理速度可以大大地提高。

2.4 霍夫投票

在获得目标模型和当前场景之间的三维对应之后,霍夫投票算法用于识别目标,同时估计其在场景中的初始姿态。此外,匹配的异常值可以通过投票过程过滤掉。霍夫投票算法分为离线和在线 2 个过程。前者产生目标模型的特征点的投票向量,后者产生目标模型和当前场景之间的对应关系的投票向量。设直矢 d_c^M 、 d_i^M 分别表示目标模型在一个全局参照系中的中心点和特征点。在离线过程中,对于目标模型的每个特征点,在局部参照系中表示的旋转和平移不变投票向量通过下式计算:

$$\mathbf{v}_{L_i}^M = \mathbf{R}_{GL_i}^M (d_c^M - d_i^M) \quad (9)$$

式中 $\mathbf{R}_{GL_i}^M$ 为从全局参照系到与特征点 d_i^M 的局部参照系相关联的变换矩阵,该变换矩阵由不变局部参照系估计算法获得^[17]。

当从描述符匹配操作获得目标模型和当前场景之间的对应关系时,激活在线过程以确定场景的全

局参照系中的第 i 个对应关系的投票向量,使得:

$$\mathbf{v}_{G_i}^S = \mathbf{R}_{L_i G}^S \mathbf{v}_{L_i}^S + \mathbf{d}_i^S \quad (10)$$

其中, $(\mathbf{d}_i^S \leftrightarrow \mathbf{d}_i^M)$ 表示场景的特征点与目标模型的特征点之间的第 i 个对应关系, $\mathbf{v}_{L_i}^S = \mathbf{v}_{L_i}^M$ 是特征点 \mathbf{d}_i^S 在局部参照系中的旋转和平移不变投票向量, $\mathbf{R}_{L_i G}^S$ 是从局部参照系到与特征向量 \mathbf{d}_i^S 相关联的全局参照系的变换矩阵。图 3 给出了基于式(9)、(10)的投票向量计算模型。由于投票向量 $\mathbf{v}_{G_i}^S$ 可以在场景的全局参照系中投票,因此场景中目标的中心点 \mathbf{d}_c^S 可以被 3D 霍夫空间中具有最大票数的单元格有效地识别。投票向量不指向同一单元的匹配点被视为异常匹配值,如图 4 中的红线所示。最后,根据识别的目标中心点 \mathbf{d}_c^S 估计目标的初始姿态,使得:

$$\begin{cases} \mathbf{R}_0^S = \mathbf{R}_{L_i G}^S \mathbf{R}_{GL_i}^M \\ \mathbf{t}_0^S = \mathbf{d}_c^S - \mathbf{d}_c^M \end{cases} \quad (11)$$

其中, \mathbf{R}_0^S 和 \mathbf{t}_0^S 是场景的全局参照系中目标的初始旋转矩阵和初始平移向量。

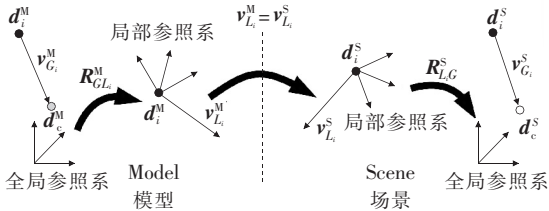


图 3 投票向量计算

Figure 3 Voting vector calculation

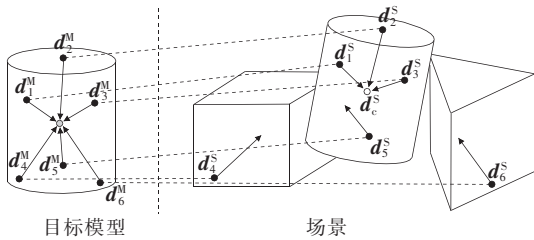


图 4 基于霍夫投票的目标识别

Figure 4 Schematic diagram of target recognition based on Hough voting

3 姿态评估

本文介绍了所提出的基于 RANSAC 算法的姿态估计方法,以改进从霍夫投票方法获得目标的初始姿态。然后,使用假设验证算法获得最优的目标

姿态估计结果。

3.1 改进的 RANSAC 算法

虽然霍夫投票算法可以提供目标的初始姿态估计,但仍需要进行优化处理,以提高姿态估计过程对匹配异常点的鲁棒性。给定目标模型和当前场景之间的 N 个对应关系为 $(\mathbf{d}_i^S \leftrightarrow \mathbf{d}_i^M)$, $i=1 \sim N$ 。 $\mathbf{T} = [\mathbf{R} \mid \mathbf{t}] \in \mathbf{R}^{3 \times 4}$ 表示与 3×3 旋转矩阵 \mathbf{R} 和 3×1 平移向量 \mathbf{t} 相关联的变换矩阵。基于对应关系之间的距离平方和 $(\mathbf{d}_i^S \leftrightarrow \mathbf{d}_i^M)$ 定义参数 $\epsilon(\mathbf{T})$,使得:

$$\epsilon(\mathbf{T}) \Big|_{(\mathbf{d}_i^S \leftrightarrow \mathbf{d}_i^M)}^N = \sum_{i=1}^N \Pi(\mathbf{T}, \mathbf{d}_i^M) - \mathbf{d}_{j_2}^{S_2} \quad (12)$$

其中, $\Pi(\mathbf{T}, \mathbf{d}) = \mathbf{R}\mathbf{d} + \mathbf{t}$ 是与变换矩阵 \mathbf{T} 相关联的 3D 点 \mathbf{d} 的刚性变换。然后,通过最小化式(12)来计算 $\mathbf{d}_i^S \leftrightarrow \mathbf{d}_i^M$ 对应关系之间的最优变换矩阵为

$$\hat{\mathbf{T}} = \underset{\mathbf{T} \in \mathbf{R}^{3 \times 4}}{\operatorname{argmin}} \epsilon(\mathbf{T}) \Big|_{(\mathbf{d}_i^S \leftrightarrow \mathbf{d}_i^M)}^N \quad (13)$$

该问题可以用 Levenberg-MarQuardt 算法求解。然而,利用式(13)进行姿态估计的方法对匹配的异常值非常敏感。为了解决这个问题,采用了 RANSAC 算法来配合式(13)的姿态估计。

$\bar{\mathbf{d}}_i^M = \prod (\mathbf{T}_0^S, \mathbf{d}_i^M)$ 表示与初始变换矩阵 $\mathbf{T}_0^S = [\mathbf{R}_0^S \mid \mathbf{t}_0^S]$ 相关联的目标模型的第 i 个变换特征点。下面列出了所提出的 RANSAC 姿态估计算法的处理步骤。

1)初始化。设置正阈值 t_{poly} 以评估多边形边长度的相似性。迭代计数器 $k=0$,最佳内围点数 $N_{\text{best}}=0$ 。设置迭代计数器为最大值 k_{max} 。

2)对于 $i=1 \sim N$,随机选择 n 个变换后的目标模型和当前场景 $(\mathbf{d}_i^S \leftrightarrow \mathbf{d}_i^M)$ 之间的对应。

3)计算 n 个采样多边形长度之间的相异向量为

$$\boldsymbol{\delta} = \left[\frac{|l_1^S - l_1^M|}{\max(l_1^S, l_1^M)} \dots \frac{|l_n^S - l_n^M|}{\max(l_n^S, l_n^M)} \right]^T \in \mathbf{R}^{n \times 1} \quad (14)$$

其中, $l_i^S = \|\mathbf{d}_j^S - \mathbf{d}_i^S\|_2$ 和 $l_i^M = \|\bar{\mathbf{d}}_j^M - \bar{\mathbf{d}}_i^M\|_2$ 分别表示场景多边形和变换的模型多边形的边长。如果 $\|\boldsymbol{\delta}\|_2 > t_{\text{poly}}$,则返回步骤 2。

4)使用 n 个对应关系 $(\mathbf{d}_i^S \leftrightarrow \bar{\mathbf{d}}_i^M)$ 估计次优变换 \mathbf{T}_1^S ,即

$$\mathbf{T}_1^S = \underset{\mathbf{T} \in \mathbf{R}^{3 \times 4}}{\operatorname{argmin}} \epsilon(\mathbf{T}) \Big|_{(\mathbf{d}_i^S \leftrightarrow \bar{\mathbf{d}}_i^M)}^N \quad (15)$$

5)对变换后的目标模型进行次优假设变换,得到假设目标模型点 $\hat{\mathbf{d}}_i^M = \prod (\hat{T}, \bar{\mathbf{d}}_i^M)$ 。

6)在 \mathbf{d}_i^S 和 $\hat{\mathbf{d}}_i^M$ 的特征点之间通过最近邻搜索找到匹配的内围点数,如果当前内围点的 N_{in} 小于最佳数 N_{best} ,则迭代计数器 $k = k + 1$ 并返回步骤 2。

7)记录当前匹配的内围点为最佳对应,并将最佳内围点数设置为 $N_{best} = N_{in}$ 。

8)更新迭代计数器的最大值为

$$k_{max} = \frac{\ln(1-p)}{\ln(1-\omega^n)} \quad (16)$$

其中, $\omega = N_{best}/N$ 为当前内围点概率, p 为期望的内围点概率, n 为步骤 2 中使用的采样数。迭代计数器 $k = k + 1$ 。如果 $k < k_{max}$,则返回步骤 2。

9)使用最佳对应 ($\mathbf{d}_i^S \leftrightarrow \hat{\mathbf{d}}_i^M$) 估计假设变换,即

$$\mathbf{T}_2^S = \underset{\mathbf{T} \in P^{3 \times 4}}{\operatorname{argmin}} (\mathbf{T}) \Bigg|_{\substack{N_{best} \\ (\mathbf{d}_i^S \leftrightarrow \mathbf{d}_i^N)}} \quad (17)$$

10)使用变换 T_0^S , T_1^S 和 T_2^S 恢复目标的最佳变换,得到当前场景中的目标的精确姿态为

$$T_{001}^S = T_2^S \circ (T_1^S \circ T_0^S) \quad (18)$$

其中, $T_2 \circ T_1 = [R_2 R_1 | t_2 + R_2 t_1]$ 。

3.2 假设验证

在霍夫投票过程中,可能在场景中检测到多于一个的目标假设。然而,并不是每个假设都对应于一个有效的目标检测结果。因此,所提出的姿态估计算法的最后一步是全局假设验证过程,该过程根据目标模型和当前场景的几何线索来评估每个目标假设。假设霍夫投票过程提供了与目标模型相关的 m 个识别假设。 $\mathbf{T}_0^S |_j$ 表示对应于第 j 个目标假设的第 j 个最优变换矩阵, $\Omega_0^S |_j = \{\mathbf{d} : \mathbf{d} \in \prod (\mathbf{T}_0^S |_j, \mathbf{d}_i^M)\}$ 是映射到当前场景中的第 j 个目标假设的点集。设 Ω^S 为当前场景的点集,然后通过权函数来测量场景点 $\mathbf{p} \in \Omega^S$ 与其最近的 $\mathbf{q} \in \Omega_0^S |_j$ 之间的局部适应度为

$$\omega_\rho(\mathbf{p}, \mathbf{q}) |_j = \left[1 - \frac{\|\mathbf{p} - \mathbf{q}\|_2}{\rho} \right]_0 \mathbf{n}_p^T \mathbf{n}_q \quad (19)$$

式中 $[x]_0$ 为限幅函数; ρ 为评估 \mathbf{p} 和 \mathbf{q} 之间距离的正阈值; \mathbf{n}_p 、 \mathbf{n}_q 分别为 \mathbf{p} 和 \mathbf{q} 的法向量。

定义一组布尔变量 $\chi^b = \{b_1, b_2, \dots, b_m\}$, 每个 $b_j \in \{0, 1\}$ 指示第 j 个识别假设是否有效。为了找

到最佳假设,采用汇总目标模型和当前场景的几何线索的全局假设验证函数 $\mathbf{I}(\chi^b) : \mathbf{B}^m \rightarrow \mathbf{R}$, 使得:

$$\mathbf{I}(\chi^b) = f_S(\chi^b) + \lambda \cdot f_M(\chi^b) \quad (20)$$

式中 λ 为正则化系数; f_M 、 f_S 为与目标模型的几何线索和由下式给出的当前场景的几何线索相关联的标量函数,即

$$f_S(\chi^b) = \sum_{\mathbf{p} \in \Omega^S} (\Lambda_{\chi^b}(\mathbf{p}) + \Gamma_{\chi^b}(\mathbf{p}) - E_{\chi^b}(\mathbf{p})) \quad (21)$$

$$f_M(\chi^b) = \sum_{j=1}^m b_j \Phi_j^M \quad (22)$$

式(21)、(22)中 $E_{\chi^b}(\mathbf{p}) = \sum_{j=1}^m b_j \omega_\rho(\mathbf{p}, \mathbf{q}) |_j$ 评估

Ω^S 和 $\Omega_0^S |_j$ 之间的适配度。 $\Lambda_{\chi^b}(\mathbf{p}) = \left[\sum_{j=1}^m \operatorname{sgn} \cdot (\omega_\rho(\mathbf{p}, \mathbf{q}) |_j) \right]_0$ 统计所有假设之间每个场景点多

次赋值的次数,其中 $\operatorname{sgn}(x)$ 为实数 x 的符号函数。

$\Gamma_{\chi^b}(\mathbf{p}) = \sum_{i=1}^m b_j r_\eta^\kappa(\mathbf{p}, \mathbf{q}) |_j$ 表示杂乱场景集合 $\Omega_0^S |_j$ 在第 j 个假设集合 Ω^S 附近的影响。

$$r_\eta^\kappa(\mathbf{p}, \mathbf{q}) |_j = \begin{cases} \kappa, & \|\mathbf{p} - \mathbf{q}\|_2 \leq \eta \text{ and } \mathbf{p} \in \Omega_0^S \\ \omega_\rho(\mathbf{p}, \mathbf{q}) |_j, & \text{otherwise} \end{cases} \quad (23)$$

式中 r_η^κ 为杂波加权函数,其中 κ 用于惩罚集合附近的未解释的场景点; η 为定义集合 Ω_0^S 的范围的正阈值; Φ_j^M 为映射到当前场景的第 j 个目标假设的异常值数量。通过约束优化过程确定最优假设,使得:

$$\hat{\chi}^b = \underset{\chi^b \in \mathbf{B}^m}{\operatorname{argmin}} S(\chi^b), \|\chi^b\|_\infty > 0 \quad (24)$$

这个问题可以用经典的模拟退火算法来解决。

4 实验结果

在使用 i5-4460 CPU 和 8 GB 系统内存的 Windows 7 平台上对所提出的识别方法进行评估。图 5 (a)为姿态估计实验中使用的水平模型,采用如图 5 (b)所示的机械臂以提供目标姿态的实际姿态信息,用于评估 3D 姿态估计结果。

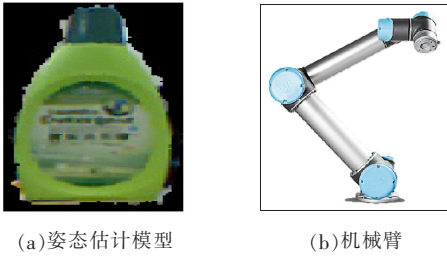


图 5 姿态估计模型和机械臂

Figure 5 Pose estimation model and robotic arm

4.1 识别结果

基于所提出方法的 3D 物体识别实验结果如图 6 所示。在该实验中,刀闸处于室外复杂环境,增加了目标识别任务难度,如图 6(a)所示。图 6(b)显示了相应的识别结果,其中绿色数据点表示从所提出的方法获得的目标识别结果。实验结果表明,即使目标被遮挡,该方法依然能够成功识别复杂环境下的目标,从而验证了该方法的目标识别性能。

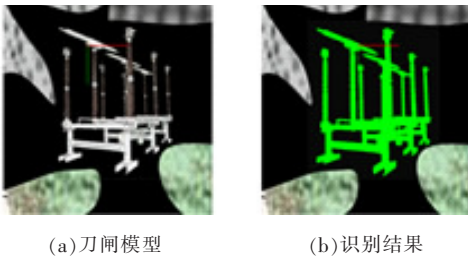


图 6 目标识别实验结果

Figure 6 Target recognition experiment results

4.2 姿态估计结果

基于所提出方法的 3D 姿态估计的实验结果如图 7 所示,其中绿色数据点也表示了该算法的目标识别结果。由图 7 可知,所提出方法成功检测出安装在机械臂末端执行器上的目标。在本实验中,根据机械臂末端执行器提供的信息记录目标的实际 3D 姿态。

目标的实际 3D 姿势和由所提出的方法得到的相应估计结果如表 1 所示,其中 t_x 、 t_y 和 t_z 分别表示目标在 x 、 y 和 z 轴上的平移量, r_x 、 r_y 和 r_z 分别表示目标关于 x 、 y 和 z 轴的欧拉角, t_x^* 、 t_y^* 、 t_z^* 和 r_x^* 、 r_y^* 、 r_z^* 是依据所提出方法获得的平移量和欧拉角。为了分析所提出方法的姿态估计误差,可以使用绝对估计误差评估系统的性能,即

$$\begin{cases} e_{\Omega}^t = |t_{\Omega} - t_{\Omega}^*| \\ e_{\Omega}^r = |r_{\Omega} - r_{\Omega}^*| \end{cases} \quad (24)$$

其中, $\Omega = \{x, y, z\}$ 为 3D 笛卡尔坐标系的 3 个轴。

所提出方法在实验中的绝对估计误差如表 2 所示。由表 2 可知, x 、 y 、 z 轴上的平均误差分别约为 0.69、0.84、0.67 cm。各轴上的最大误差均小于 2.0 cm,验证了该算法对目标平移估计的准确性。而 z 轴上的旋转估计误差最大,平均约为 3.53° 。相比之下, x 、 y 轴的平均绝对旋转误差分别约为 1.42° 和 3.27° 。上述实验结果表明,本文提出的方法能够提供目标的精确 3D 姿态估计结果。

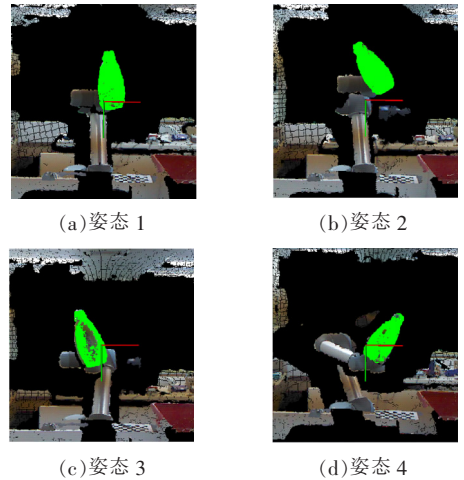


图 7 姿态评估实验结果

Figure 7 Pose assessment experiment results

表 1 姿态估计结果对比

Table 1 Comparison of pose estimation results

方法	实际姿态						评估姿态					
	t_x	t_y	t_z	r_x	r_y	r_z	t_x^*	t_y^*	t_z^*	r_x^*	r_y^*	r_z^*
图 7(a)	5	0	55	10	-10	-10	4.648	0.313	54.701	12.849	-6.411	-9.868
图 7(b)	5	-5	60	15	0	-30	3.449	-5.996	59.160	14.677	7.293	-34.556
图 7(c)	-5	5	60	-5	15	-15	-5.821	3.082	60.303	-4.827	14.265	-12.258
图 7(d)	10	10	70	-15	-10	30	9.941	10.171	71.268	-12.646	-8.534	23.301

表 2 姿态估计结果误差

方法	误差					
	e_x^t	e_y^t	e_z^t	e_x^r	e_y^r	e_z^r
图 7(a)	0.352	0.313	0.299	2.849	3.589	0.132
图 7(b)	1.551	0.996	0.840	0.323	7.293	4.556
图 7(c)	0.821	1.918	0.303	0.173	0.735	2.742
图 7(d)	0.059	0.171	1.268	2.354	1.466	6.699
平均	0.695	0.849	0.677	1.424	3.271	3.532

4.3 计算效率

所提出方法的每个步骤的平均处理时间如表 3 所示,可知所提出方法的总处理时间取决于在场景

表 3 所提出方法的平均处理时间

特征数	处理时间						总时间
	场景分割	特征描述	描述符匹配	霍夫投票	RANSAC 算法	假设验证	
1 000	0.185	0.358	0.863	0.641	0.052	0.062	2.161
3 000	0.201	0.524	1.857	0.625	0.190	0.093	3.490

表 4 所提出方法和 $k-d$ 树方法处理时间对比Table 4 Comparison of processing time between the proposed method and the $k-d$ tree method

特征数	k-d 树方法	所提出方法	变化量/%
<1 000	1.2	0.8	-33.3
>4 000	6.2	3.4	-45.2

5 结语

本文提出了一种新的基于点云特征提取的三维图像识别方法。在描述符匹配方面,提出了一种二阶段匹配算法,大大地提高了特征匹配的速度,特别是在特征数量急剧增加的情况下;在目标识别方面,将几何形状特征与颜色纹理相结合,有效地实现了三维目标检测与识别。实验结果表明,该方法不仅能够复杂的场景中成功地识别出目标,而且能够提供准确的姿态估计结果。3 个轴的平移和旋转估计误差分别小于 2.0 cm 和 5.0°。因此,实验结果验证了该方法的性能。

参考文献:

[1] 翟俊海,赵文秀,王熙照. 图像特征提取研究[J]. 河北大

点云中检测到的特征的数量,当特征的数量增加到 4 000 个关键点时,处理时间从 2.16 s 增加到 3.49 s。此外,描述符匹配阶段需要的处理时间最长,特别是当特征数目变大时,这一结果突出了所提出的描述符匹配方法的重要性。基于 $k-d$ 树的 NNS 方法和所提出的两阶段匹配方法之间的处理时间比较如表 4 所示。当特征个数小于 1 000 时,该方法可以使 $k-d$ 树方法的整体描述符匹配时间减少约 33.3%。并且,当特征数大于 4 000 时,该方法显著减少了 $k-d$ 树方法的描述符匹配时间,平均可达 45.2%。因此,当特征数目快速增加时,所提出的方法效率更高。

学学报(自然科学版),2009,29(1):106-112.

ZHAI Junhai,ZHAO Wenxiu,WANG Xizhao. Research on the image feature extraction[J]. Journal of Hebei University (Natural Science Edition), 2009, 29(1): 106-112.

[2] 李旭冬,叶茂,李涛. 基于卷积神经网络的目标检测研究综述[J]. 计算机应用研究,2017,34(10):2881-2886+2891.

LI Xudong, YE Mao, LI Tao. Review of object detection based on convolutional neural networks[J]. Application Research of Computers, 2017, 34(10): 2881-2886 + 2891.

[3] 郑远攀,李广阳,李晔. 深度学习在图像识别中的应用研究综述[J]. 计算机工程与应用,2019,55(12):20-36.

ZHENG Yuanpan, LI Guangyang, LI Ye. Survey of application of deep learning in image recognition[J]. Computer Engineering and Applications, 2019, 55(12): 20-36.

[4] OZAY M, WALAS K, LEONARDIS A. A hierarchical approach for joint multi-view target pose estimation and categorization[C]//2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 2014:5480-5487.

- [5] ZHANG A L, ZHU E G, DONG F S. Viewpoint calibration method based on point features for point cloud fusion[C]//2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 2017:2224-2228.
- [6] JEONG D, CHOO S, SEO W. Regional deep feature aggregation for image retrieval[C]//2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 2017:1737-1741.
- [7] RUSU R B, BLODOW N, BEETE M. Fast point feature histograms (FPFH) for 3D registration[C]//IEEE International Conference on Robotics and Automation, Kobe, Japan, 2016:3212-3217.
- [8] TANG K, SONG P, CHEN X. 3D target recognition in cluttered scenes with robust shape description and correspondence selection[J]. IEEE Access, 2017, 5(6):1833-1845.
- [9] ZHANG Z, ZHAO M, CHOW T W S. Graph based constrained semi-supervised learning framework via label propagation over adaptive neighborhood [J]. IEEE Transactions on Knowledge and Data Engineering, 2015, 27(9):2362-2376.
- [10] ZHEN M, WANG W M, WANG R G. Signature of unique angles Histograms for 3D data description[C]//2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Turin, Italy, 2015:384-390.
- [11] JIANG X Y, LU T, TIAN X. Robust feature matching for remote sensing image registration via linear adaptive filtering[J]. IEEE Transactions on Geoscience and Remote Sensing, 59(2), 2021:1577-1591.
- [12] MA J Y, QIU W C, ZHAO J, et al. Robust L2E estimation of transformation for non-rigid registration[J]. IEEE Transactions on Signal Processing, 63(5), 2015:1115-1129.
- [13] CHEN S L, HONG J, LIU X N. A framework for 3D target detection and pose estimation in unstructured environment using single shot detector and refined line MOD template matching[C]//2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Zaragoza, Spain, 2019:499-504.
- [14] CIVIR C, TOPAL. Robust landmark selection for 3D face pose estimation[C]//2019 27th Signal Processing and Communications Applications Conference (SIU), Sivas, Turkey, 2019:1354-1361.
- [15] HOLZER S, RUSU S B, M. Dixon. Adaptive neighborhood selection for real-time surface normal estimation from organized point cloud data using integral images [C]//2012 IEEE International Conference on Intelligent Robots and Systems, Vilamoura-Algarve, Portugal, 2012:2684-2689.
- [16] KANG S, KIM J, YUAN J. A fast region expansion labeling of connected components in binary image[C]//The 18th IEEE International Symposium on Consumer Electronics (ISCE 2014), Jeju, Korea, 2014:345-351.
- [17] PETRELLI A, STEFANO L D. On the repeatability of the local reference frame for partial shape matching [C]//2011 International Conference on Computer Vision, Barcelona, Spain, 2011:2244-2251.