

面向三维特征描述子的自适应二进制简化方法

刘双元^{1,2}, 郑王里³, 林云汉^{1,2*}

(1. 武汉科技大学 计算机科学与技术学院, 武汉 430081;

2. 智能信息处理与实时工业系统湖北省重点实验室(武汉科技大学), 武汉 430081;

3. 国网电力科学研究院有限公司, 南京 211106)

(* 通信作者电子邮箱 yhlin@wust.edu.cn)

摘要: 在三维(3D)局部特征描述子研究中, 准确度、匹配时间以及内存消耗存在此消彼长的问题。针对上述问题, 基于统计理论中的标准差原理提出一种面向3D特征描述子的自适应二进制简化方法。首先, 通过改变简化模型中二值化单元长度和标准差个数来生成不同的二进制特征描述子; 然后, 将它们应用到当前被广泛使用的基于签名的方向直方图(SHOT)描述子中, 并通过实验确定最优的二值化单元长度和标准差个数的组合; 最后, 将最优组合下的简化描述子命名为SD-SHOT。实验结果表明, 与未进行简化的SHOT描述子相比, SD-SHOT在关键点匹配时间上减少为原来的1/15, 内存占有率降低为原来的1/32; 与现有主流简化方法如B-SHOT等相比, SD-SHOT的性能达到了综合最优水平。此外, 在由五种不同类别的物体构成的实际机器人分拣场景中验证了所提方法的有效性。

关键词: 自适应二进制方法; 三维特征描述子; 三维物体识别; 点云; 物体分拣

中图分类号: TP391.4 **文献标志码:** A

Adaptive binary simplification method for 3D feature descriptor

LIU Shuangyuan^{1,2}, ZHENG Wangli³, LIN Yunhan^{1,2*}

(1. School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan Hubei 430065, China;

2. Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System
(Wuhan University of Science and Technology), Wuhan Hubei 430065, China;

3. State Grid Electric Power Research Institute Company Limited, Nanjing Jiangsu 211106, China)

Abstract: In the study of 3-Dimensional (3D) local feature descriptor, it is difficult to strike a balance among accuracy, matching time and memory consumption. To solve this problem, an adaptive binary simplification method for 3D feature descriptor was proposed based on the standard deviation principle in statistical theory. First, different binary feature descriptors were generated by changing the binarization unit length and the number of standard deviations in the simplification model, which were applied into the currently widely used Signature of Histogram of Orientations (SHOT) descriptor, and the optimal combination of binarization unit length and the number of standard deviations was determined by experiments. Finally, the simplified descriptor under the optimal combination was named Standard Deviation feature descriptor for Signature of Histogram of Orientations (SD-SHOT). Experimental results show that compared with the SHOT descriptor without simplification, SD-SHOT reduces the key point matching time to 1/15 times and the memory occupancy to 1/32 times of SHOT; compared with the existing mainstream simplification methods such as Binary Feature Descriptor for Signature of Histogram of Orientations (B-SHOT), SD-SHOT has the optimal comprehensive performance. In addition, the validity of the proposed method is verified in the actual robot sorting scene consisting of five different categories of objects.

Key words: adaptive binary method; 3-Dimensional (3D) feature descriptor; 3-Dimensional (3D) object recognition; point cloud; object sorting

0 引言

环境视觉感知可分为二维(2 Dimension, 2D)感知和三维(3 Dimension, 3D)感知。3D的感知可构建三维世界模型, 确定场景中可见物体的3D位姿、名称、大小、形状和颜色等信息, 更符合人类对世界的感知。此外, 在光照条件不好的时候, 如果采用红外、激光技术提供的3D数据可以获得比2D图

像或者人眼更好的结果。当前3D环境感知算法可分为基于深度学习的方法和基于手工计算的方法。基于深度学习的方法虽然近几年来取得了一定的突破, 然而, 该技术是基于一个封闭、静态环境的, 即训练数据和测试数据是来自相同的标记和特征空间; 同时, 其识别正确率的高低依赖于带注释/标记的训练集的多少。所以, 基于深度学习的方法大多数仅应用于离线数据处理, 暂时还不适用于机器人实际在线的工业应

收稿日期: 2020-09-27; 修回日期: 2020-12-22; 录用日期: 2020-12-24。

基金项目: 湖北省自然科学基金青年项目(2020CFB116); 湖北省教育厅科学研究计划中青年人才项目(Q20191108)。

作者简介: 刘双元(1995—), 男, 山西忻州人, 硕士研究生, 主要研究方向: 机器人感知、三维视觉; 郑王里(1985—), 男, 福建仙游人, 工程师, 硕士, 主要研究方向: 电力通信设备检测; 林云汉(1985—), 男, 福建仙游人, 讲师, 博士, 机器人感知、机器人推理与控制。

用场景^[1]。然而,基于手工计算的方法通过人为设计合理的特征描述方法来表征3D点云数据,之后通过场景和模型之间的特征匹配实现正确的识别,只要能够实现有效的特征表征,即可实现较好的在线环境感知。所以,目前基于手工计算的方法在3D视觉领域的研究和应用仍然是研究的热点。根据数据特征表征方式的不同,基于手工计算的3D环境感知又可

分为基于全局特征描述子的感知方法和基于局部特征描述子的感知方法。基于特征局部描述子的感知方法因为它在遮挡、重叠、背景杂乱以及视角变化等场景中具有较好的鲁棒性而被广泛地研究。基于特征局部描述子的总体流程如图1所示,主要包括物体模型库建立、物体识别与位姿估计和语义地图生成三个部分。

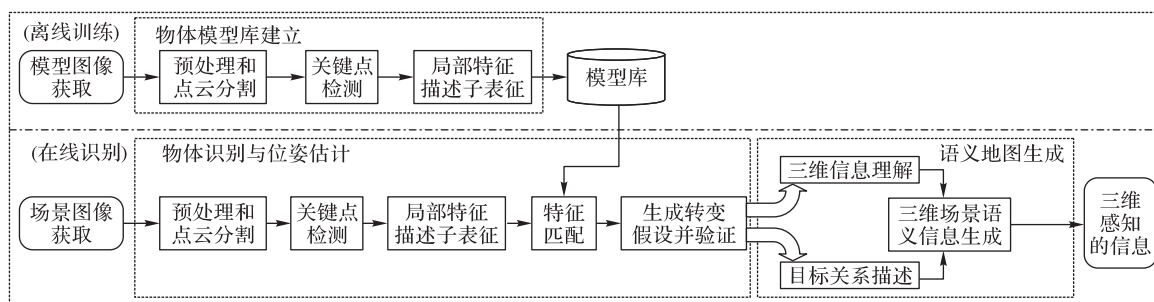


图1 基于局部特征描述子的3D视觉感知流程

Fig. 1 3D visual perception process based on local feature descriptors

然而,现有的3D局部特征描述子算法在保证良好识别正确率(较好的描述性和鲁棒性)的同时,存在匹配耗时长、内存占用高的缺点。针对该问题,本文研究面向3D点云快速特征匹配的自适应二进制局部特征描述子,实现3D环境实时感知与精确场景理解^[2]。从点云数据的离散性和3D特征描述子的设计原理出发,分析3D特征描述子的特点,对3D特征描述子进行二进制量化,量化后的特征描述子性能优于现有算法,在物体识别精度和特征匹配效率两方面能达到目前综合最优水平。

1 相关工作

3D特征描述子表征点云的几何特征,在点云的识别、曲面重建、配准、重采样等过程中,都需要采用特征描述子对点云数据进行特征的描述。特征描述子从尺度上面可分为:全局特征描述子和局部特征描述子。

全局特征描述子如视点特征直方图(Viewpoint Feature Histogram, VFH)、聚类视点特征直方图(Clustered Viewpoint Feature Histogram, CVFH)、定向独特可重复的聚类视点直方图(Oriented, Unique and Repeatable Clustered Viewpoint Feature Histogram, OUR-CVFH)、全局一致的空间分布(Globally Aligned Spatial Distribution, GASD)特征描述子、正交视点特征直方图(Orthogonal Viewpoint Feature Histogram, OVFH)特征描述子^[3]、部分视点特征直方图(Partition Viewpoint Feature Histogram, PVFH)特征描述子等都是在整个场景中分割出每个物体的聚类,然后采用全局特征描述子对每个物体聚类点云的几何信息进行描述。然而,当场景中背景杂乱或物体出现遮挡时,分割出的物体聚类可能不正确,表征的实际场景的物体点云特征信息存在失真的问题,最终导致实际模型与场景特征描述匹配错误,识别效果较差。

在实际的3D物体感知应用中,物体间遮挡和物体背景杂乱的情况经常存在。与基于全局的特征描述子感知方法不同,基于局部特征描述子的视觉感知首先计算物体的关键点,然后通过局部特征描述子来描述关键点邻域内点云的几何特征,如点特征直方图(Point Feature Histogram, PFH)^[4]、快速点

特征直方图(Fast Point Feature Histograms, FPFH)^[5]、基于签名的方向直方图(Signature of Histogram of Orientations, SHOT)^[6]、局部表面(Local Surface Descriptor, TriSI)^[7]、旋转投影统计(Rotational Projection Statistics, RoPS)^[8]、距离直方图(Histogram of Distances, HoD)^[9]、以及2016年和2017年,Yang等^[10-12]提出的局部特征统计直方图(Local Feature Statistics Histogram, LFSH)、多属性统计直方图(Multi-attribute Statistics Histograms, MaSH)和三重正交局部深度图像表征(Triple Orthogonal Local Depth Images, TOLDI)特征描述子;2019年Zhao等^[13]提出的细分空间上偏差角统计(Statistics of Deviation Angles on Subdivided Space, SDASS),以及Zhao等^[14]在2020年初新发表的论文中提出的分区本地特征统计(Divisional Local Feature Statistics, DLFS)。由于基于局部的特征描述子关注的是物体的关键点和关键点周围的点云信息,所以无需对整个物体的点云进行描述。基于局部特征描述子的方法一方面提高了物体识别的实时性,另一方面也解决了背景杂乱和物体间存在遮挡的情况下物体识别正确率低的问题^[8]。

文献[14-16]中对比分析了十来种当前主流的局部特征描述子的性能,在多个不同的数据集下进行了比较和评估,从评估的结果可以看出:在物体识别过程中,特征描述子算法提取到物体的点云的几何特征越多,识别效果越佳,但是处理时间越长;同时,不同算法对于不同的环境特点表现出了不同的识别性能。

表1列举了12种主流局部特征描述子的维度对比。一方面,这些特征描述子每个维度的值都以浮点值(real-valued)形式存储,维度越少特征匹配的速度就越快;另一方面,这些算法的提出都围绕着特征描述子的描述性、鲁棒性以及匹配耗时、内存占用等评价指标展开评估。如何在保证特征描述子良好识别正确率(较好的描述性和鲁棒性)的同时提高特征匹配效率并降低内存占用是3D视觉领域的核心和瓶颈问题。所以,近几年来探索如何提高特征匹配效率和降低内存占用的研究成了很多研究者感兴趣的热点。

表 1 不同文献中的特征描述子占用内存比较

Tab. 1 Comparison of memory consumption by feature descriptors from different literatures

序号	特征描述子	文献	发表的年份	维度
1	PFH	[4]	2008	125
2	FPFH	[5]	2009	33
3	SHOT	[6]	2010	352
4	TriSI	[7]	2013	675
5	RoPS	[8]	2013	135
6	HoD	[9]	2016	240
7	LFSH	[10]	2016	30
8	MaSH	[11]	2017	140
9	TOLDI	[12]	2017	1 200
10	SDASS	[13]	2019	345
11	DLFS	[14]	2020	240
12	CDLFS	[14]	2020	78

目前,实现特征匹配效率提高和内存占用降低的主要方法有三种:1)压缩已有特征描述子维度;2)设计全新的二进制特征描述子;3)二进制简化已有特征描述子。本文拟在已有特征描述子上进行有效自适应二进制简化,从而在不失精度的情况下提高匹配效率并降低内存占有率。

二进制简化已有特征描述子的方法是设计一种二进制的简化方法并将其应用于某种已经存在的特征描述子中,将原来每一维度的向量转化为二进制0或1,并在保证特征描述子的描述性和鲁棒性相当的情况下,大幅减少内存占用,并提高3D物体匹配/识别速度。在3D图像处理领域,应用二进制方法简化已有特征描述子的研究可追溯到2015年,新加坡南洋理工大学和香港科技大学的Prakhya等在机器人顶级国际会议IROS上提出了向量直接进行二进制简化的方法,该方法是一种通用的方法,可以应用于不同的特征描述子的二进制简化中。当时作者将该方法应用于传统SHOT特征描述子^[6]的二值简化中提出了B-SHOT^[17]:首先将SHOT描述子中每四个浮点型数据作为一个简化的单元,根据定义的5条规则将SHOT中的浮点型数据分别转化为二进制0和1,降低了物体表征时内存的占用,提高了匹配运算的效率,实现了3D点云快速特征匹配。2017年,该作者将该二值化的思想又应用到另外两种特征描述子RoPS^[8]和FPFH^[5]中,建立了B-RoPS和B-FPFH^[18],并通过一些扩展实验将B-SHOT、B-RoPS和B-FPFH的性能进行了对比分析。实验结果表明,B-SHOT的性能优于其他两种二值化特征描述子,该算法在3D点云关键点匹配上具有更加快速和高效的优点,而且它的内存占用降低到了SHOT的1/32,同时关键点匹配速度提高了6倍。然而,B-SHOT存在选择的简化单元不连续的问题和严重的信息丢失问题。Kechagias-Stamatis等^[19]提出了二进制距离直方图(Binary-Histogram of Distances, B-HoD)特征描述子,在已有的三维特征描述子算法HoD^[9]的基础上,通过提出一种二进制简化方法和预处理步骤,使得简化后的内存占用降低到原来的2/9,关键点匹配时间降低到原来的1/75;但是该二进制简化方法仅针对HoD特征描述子,不是一种通用的简化方法。通用的简化方法除了文献[17-18]中提出的方法之外,2017年Prakhya等^[20]采用格子量化(lattice quantization)的理论提出了

一种压缩3D特征描述子的二进制简化方法,并将这种通用的方法应用于SHOT、RoPS和FPFH的简化中产生了三种简化描述子Type-SHOT、Type-RoPS和Type-FPFH。实验结果表明,Type-SHOT效果优于Type-RoPS和Type-FPFH。与B-SHOT^[18]相比,Type-SHOT的识别正确率优于B-SHOT,但是其关键点匹配时间却比B-SHOT长,而且该方法生成简化特征描述子的时间很长,导致综合物体识别时间不降反升。本文作者团队在前期也针对通用的二进制简化方法进行了一些尝试,于2018年和2019年分别提出了两种通用的特征描述子简化方法,并将它们应用于SHOT,给出了新的描述算子CI-SHOT^[21]和GRAY-SHOT^[22]。实验结果表明,CI-SHOT和GRAY-SHOT的识别正确率也优于B-SHOT,但是它们的匹配时间略长于B-SHOT。从当前各种算法的比较来看,此类方法目前的瓶颈是:物体识别正确率与特征匹配效率在各种二进制简化算法中存在此消彼长的关系,目前还没有一种简化方法能够实现综合性能最优。

2 本文方法

本文提出一种基于统计方法的自适应二进制简化方法。首先对实际数据集/场景进行统计学分析,计算数据分布的特点并计算标准差;接下来通过实际值与标准差之间的关系进行简化编码;最后将编码后的所有二进制值联结起来形成对关键点邻域信息的二进制表征(二进制特征描述子)。从相关工作中的文献分析可知,文献[18]中提出的二进制简化方法B-SHOT是目前所有二进制通用简化算法中匹配速度最快的。该方法的二进制简化规则如下:

将SHOT特征描述子的352维特征向量 $S(S_0, S_1, S_2, \dots, S_{351})$,以每4个维的向量为一个简化单元进行二进制简化,在每一个简化单元中,令 $S_{\text{sum}} = S_0 + S_1 + S_2 + S_3$,共定义如下五种二值化规则:

规则1 若四个 S_i 的值都为0,则全部编码为0;

规则2 若某一个值 S_i 超过 S_{sum} 的90%,则其编码为1,其余为0;

规则3 若两个值的和超过 S_{sum} 的90%,则这两个值都编码为1,其余为0;

规则4 若三个值的和超过 S_{sum} 的90%,则这三个值都编码为1,其余为0;

规则5 若规则1、2、3、4都不满足,则全部编码为1。

然而,B-SHOT存在如下两个主要问题:

问题一 选择的简化单元不连续。上述简化方法以4个向量值(4个维度)为一个基本单元进行简化,但这种处理方法并没有合理的依据。从SHOT特征描述子的构建原理可知:SHOT是将关键点某一邻域内的点云通过横向、纵向和径向拆分成的空间网格(拆分成了32个网格)以及各个网格中通过计算各点云法线与关键点法线夹角的余弦值(拆分成了11个值)组成,合计构成 $32 \times 11 = 352$ 个维度,通过统计分布在这352个维度中的点云的数量并归一化处理后得到352维的统计直方图。B-SHOT将每4个维度的向量作为一个简化单元,而11并不能被4整除,所以存在一些跨越空间网格的分

组,必然导致信息的不连续。

问题二 严重的信息丢失。上述规则以 S_{sum} 的90%为判定标准来进行二进制简化,然而对于适用于规则3和4的情况时,有可能存在严重信息丢失的问题。如:SHOT特征描述子中某简化单元的四个维度的值为: $\{S_0, S_1, S_2, S_3\} = \{0.0530, 0.0040, 0.0039, 0.0011\}$,由简化的规则可知,本组数据属于“规则3”的情况,二进制简化后对应的值为: $\{B_0, B_1, B_2, B_3\} = \{1, 1, 0, 0\}$,即 B_0 和 B_1 的值都为1,而实际上 S_0 和 S_1 的值相差很大却都被简化成了1,而 S_1 和 S_2 的值基本相等却一个是1,另一个是0。故存在严重信息丢失的问题。

针对以上问题,本文在简化算法中引入简化单元参数 L ,使不同的特征描述子可根据其特点选择不同的 L ,以解决简化单元不连续的问题。另一方面,从点云数据的离散性这一基本特性出发,采用标准差的理论设计简化规则,使得在简化过程中编码能够根据简化单元中数据的概率统计情况自适应调整,以解决简化过程中信息严重丢失的问题。

3D感知的数据是以点云格式存储的,与二维图像数据以规则的像素点存储不同,点云数据的分布最显著的特点是它的无序性和不规则性(即每个点云的分布是离散的)。在统计学中,标准差(Standard Deviation, SD)是用于量化一组数据值的变化或离散程度的度量。对于离散数据的处理,如果能够从标准差理论的角度出发进行点云数据分析和简化算法设计将会是一个较好的切入点,即:对点云数据集,首先计算该集合中数据的离散程度(计算出其平均值和标准差),再根据标准差所表示的数据的分布概率情况设计简化算法。

所以,在标准差理论的启发下,本文提出了一种基于标准差的特征描述子简化方法,简化模型定义如下:

$$SD_i = f(D_i, L, n) \quad (1)$$

其中: $f()$ 是规则函数,用于将特征描述子的实际值转换为二进制值; SD_i 代表二进制简化的结果; D_i 表示原始特征描述子中的第 i 维向量的实际值; L 是二进制简化的单位长度; n 为标准差的个数,其中,标准差的值由单位长度 L 内点云数据集的值计算得到,不同的集合将得到不同的结果(存在自适应性)。所以,本模型可以通过选择不同的 L 和 n 来产生不同的简化特征描述子。简化公式如式(2):首先,计算每个单元 L 中 D_i 的平均值($MEAN$)和标准差(σ);然后将所有 D_i 与该简化单元中的 $MEAN + n \times \sigma$ 的值进行比较,如果 $D_i > MEAN + n \times \sigma$,则将 D_i 的值编码为1,否则编码为0。

$$SD_i = \begin{cases} 1, & D_i > MEAN + n \times \sigma \\ 0, & \text{其他} \end{cases} \quad (2)$$

其中:

$$MEAN = \frac{1}{L} \sum_{i=0}^{L-1} D_i \quad (3)$$

$$\sigma = \sqrt{\frac{1}{L} \sum_{i=0}^{L-1} (D_i - MEAN)^2} \quad (4)$$

简化方法步骤如下:

1)分析目标特征描述子 $D(D_0, D_1, D_2, \dots, D_{m-1})$ 的特性,并选择二进制简化单元 L ;

2)计算 L 的平均值和标准差;

3)选择标准差的个数 n ,遍历特征描述子 D_i 的每个实数值,根据规则函数将 D_i 编码为 SD_i ;

4)联结每个编码单元形成完整的二进制特征描述子。

经过二值简化后,特征描述子的实际值将转换为仅包含0和1的二进制值。因此,本文将采用汉明距离代替欧氏距离进行特征匹配。长度相等的两个字符串之间的汉明距离是对应符号不同的值的个数。

算法伪代码如下所示:

算法 基于标准差的特征描述子二进制通用简化方法

```

1) input  $D_i$  其中  $0 \leq i \leq m-1$ 
   //输入特征描述子的值  $D\{D_0, D_1, D_2, \dots, D_{m-1}\}$ 
2) input  $L$ 
   //分析输入特征描述子的特性,选择二进制简化单元  $L$ 
3) input  $n$ 
   //选择第二个简化参数:标准差的个数  $n$ 
4) for  $j$  from 0 to  $m/L-1$   $j++$ 
5)   for  $i$  from  $j \times L$  to  $(j+1) \times L-1$   $i++$ 
6)      $SUM \leftarrow SUM + D_i$ 
7)   end for
8)    $MEAN \leftarrow SUM/L$ 
9)   for  $i$  from  $j \times L$  to  $(j+1) \times L-1$ 
10)     $VARIANCE \leftarrow VARIANCE + (D_i - MEAN) \times (D_i - MEAN)$ 
11)  end for
12)   $VARIANCE \leftarrow VARIANCE/L$ 
13)   $\sigma \leftarrow \sqrt{VARIANCE}$ 
14)  for  $i$  from  $j \times L$  to  $(j+1) \times L-1$   $i++$ 
15)    if  $D_i > MEAN + n \times \sigma$ 
16)       $SD_i \leftarrow 1$ 
17)    else
18)       $SD_i \leftarrow 0$ 
19)    end if
20)  end for
21) end for
22) output  $SD$ 
   //输出二进制特征描述  $SD\{SD_0, SD_1, SD_2, \dots, SD_{m-1}\}$ 

```

3 实验与分析

首先通过文献分析选择一种浮点型特征描述子,将本文提出的特征描述子二进制简化方法应用其中,然后选择标准数据集从多个评价指标上对简化后的特征描述子进行验证、评估和对比分析。

3.1 待简化特征描述子、数据集以及评价指标的选择

2017年Prakhya等在文献[18]中将他们提出的二进制简化方法应用于实际的Kinect数据集中进行测试,得到了这样一个结论:在Kinect数据集中的性能顺序为:SHOT>B-SHOT>FPFH>RoPS>B-RoPS>B-FPFH。即:SHOT的综合识别和匹配效果优于RoPS和FPFH的同时,B-SHOT的综合识别和匹配效果也优于B-RoPS和B-FPFH。考虑到本文算法的应用领域将是机器人智能分拣系统,系统采用的是Kinect进行三维物体识别,并以识别结果为基础生成三维语义地图,作为整个“人-机-环境”交互的推理机制^[23]的输入,所以本文对于局部特征描述子的简化将重点考察算法在实际Kinect数据集中的性能,并最终将算法集成于实际的机器人自动作业系统中,实现

机器人的智能推理与自动作业^[24]。所以本文选择 SHOT 为实验对象,将本文提出的自适应简化方法应用于 SHOT 中,并在 Kinect 数据集 Bologna Kinect 数据集^[25]中进行算法性能评估。

Bologna Kinect 数据集是 Bologna 大学 Tombari 等创建的 Kinect 数据集,数据集里每个模型的点云图都是通过不同的视角对实际的目标物体进行扫描得到的;对于一个场景中的多个物体,位置摆放都是随机的,这样就构成了多个不同的场景,而且每个场景的 3D 转换信息都包含其中。此数据集一共包含 6 个模型、46 个场景。

本文采用变换矩阵差(T_{diff})来判定在不同参数设定下简化描述子的性能的评估;采用精确召回曲线(Precision-Recall Curve, PRC)^[21]来判定算法的整体性能,通过计算关键点匹配的平均耗时来分析关键点匹配的耗时情况;通过对简化前后特征描述子的内存占用比的比较展示简化前后内存占用的情况。

3.2 实验设定与参数确定

1) 计算机配置:所有实验都在同一台计算机上完成,参数: Intel Core i7 CPU 7700K@ 4.20 GHz×832 GB RAM 和 UBUNTU 16.04。

2) 实验方法:本小节的实验通过三个步骤完成:a)提取关键点;b)对关键点周围的信息进行描述;c)找到场景和模型之间的匹配点和三维转换矩阵(Translate and Rotate, TR)。在实验参数设置中,对于关键点提取的半径(R_{kp})和特征描述半径(R_{id})设置为 $R_{\text{kp}}=0.03$ m, $R_{\text{id}}=0.10$ m,实验过程中, R_{kp} 和 R_{id} 均保持不变。

3) 平均值和标准差计算:根据第 2 章中对本文算法的描述,在进行简化之前,需要对目标数据集/场景的点云信息用待简化的描述子表征后统计其平均值和计算标准差。本文选择的是 Bologna Kinect 数据集,对其按照等式(2)和(3)计算其平均值和标准差。

4) 简化单元 L 、标准差个数 n 的选择与评估:针对 SHOT 特征描述子,由于 SHOT 的 352 维向量是由空间的 32 个网格(spatial grid)以及各个网格中的 11 个 bin 组成的^[6],所以考虑以每个空间网格为单位(即将每 11 维设定为一个简化单元);或者直接以整体的 352 维为简化单元对 SHOT 进行简化,这样可以满足信息的连续性。对于标准差的个数 n ,针对点云数据的离散程度,计算数据的方差,进而对标准差的个数进行测试,最终选择 0.5、1 和 2 三种取值分别进行测试,以验证不同参数组合验证简化效果。

对于不同参数的组合,本文以参数变换矩阵差(T_{diff})为指标来进行简化描述子的性能的评估。对于 T_{diff} 表达的是实际描述子表征得到的旋转矩阵(T_{R_g})和数据集提供的基准(ground truth)旋转矩阵(T_{GT_g})之间的差异,值应该是越小越好,计算公式如式(5)。

$$T_{\text{diff}} = \sqrt{\sum_{i=1}^n \sum_{j=1}^n (T_{R_{g_i}} - T_{GT_{g_j}})^2} \quad (5)$$

测试结果如表 2 显示, $L = 352$, $n = 1$ 时,本文提出的简化方法的 T_{diff} 值与 SHOT 最相近,因此,本文将 STD-SHOT(352, 1)

命名为 SD-SHOT,而且在后续的实验中将 SD-SHOT 与其他简化描述子进行对比分析。

3.3 对比分析实验

本节将本文提出的 SD-SHOT 与 B-SHOT、GRAY-SHOT、CI-SHOT 以及原始未简化描述子 SHOT 进行性能上的比较,比较的指标包含 PRC、关键点特征匹配的耗时以及内存占用率。

在计算不同简化描述子的 precision 和 recall 时,通过改变参数 τ 的取值来验证不同算法的准确度。计算过程中:当 $\tau > 0.5$ 时,数据不再发生变化,因此本文设置 τ 的区间为 0~0.5,步长为 0.05,共 10 组数据进行测试,其余参数如 3.2 节所示保持不变。实验结果如图 2 显示,SD-SHOT 的整体识别效果均优于 CI-SHOT、GRAY-SHOT 和 B-SHOT,也更加接近原始未简化的描述子 SHOT。需要注意的是在 τ 的值取值为 0.05、0.10 以及 0.15 时,本文删除了一些明显错误匹配点并对 NN-Matching 和 RANSAC 后匹配点的数量为 0 的点进行了处理(1-precision 和 recall 的定义中分母不能为零,所以本文将这些点的 1-precision 和 recall 的值统一设置为 0)。

表 2 不同参数组合的简化描述子 T_{diff} 对比

Tab. 2 Comparison of simplified descriptor T_{diff} of different parameter combinations

特征描述子	简化单元长度 L	标准差个数 n	T_{diff}
SHOT			17.15
B-SHOT	4		48.85
STD-SHOT	11	0.5	34.52
STD-SHOT	11	1	25.69
STD-SHOT	11	2	46.14
STD-SHOT	352	0.5	20.92
STD-SHOT	352	1	20.15
STD-SHOT	352	2	23.77

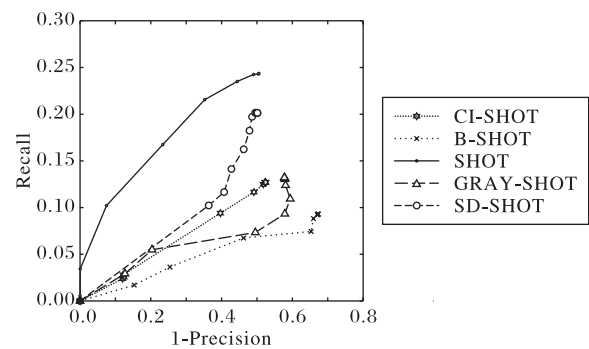


图 2 不同简化描述子的 PRC

Fig. 2 PRC of different simplified descriptors

针对特征点匹配的平均耗时及内存占用率的比较,如表 3 和表 4 所示。平均耗时是从特征描述子的简化开始直到 RANSAC 结束,其中每个步骤的平均计算时间是来自数据集的 46 个“场景-模型”对的平均时间,所有参数中除简化方案外均保持一致。从表 3 可知,本文提出的 SD-SHOT 时间减少率优于所有的简化特征描述子,也优于 FPFH 局部特征描述子。这里没有统计 SHOT 描述子、FPFH 描述子和 SD-SHOT 以及其他二进制简化描述子的建立时间,仅比较描述子特征简化时间和实际匹配时间总和。从表 3 中可以看出,对于 SHOT 原始描述符,在使用二进制进行简化后匹配时间都有大

幅度的降低。对于内存占用率的比较,SD-SHOT采用一位二进制表示每一维SHOT特征描述子的浮点值,所以采用了352位表示352维SHOT浮点值向量,由于SHOT是352维浮点数,

因此它需要1408字节(11264位)的二进制数据。从表4可知,表现最优的特征描述子是SD-SHOT和B-SHOT,它们的内存占用率降低到了原来SHOT的3.13%。

表3 关键点特征匹配的平均耗时比较

Tab. 3 Average time consumption comparison of key point feature matching

描述子名称	简化耗时/s	NN-Matching/s	RANSAC/s	总时间/s	各描述子总时间与SHOT总时间的比值/%
SD-SHOT	0.001 462	0.003 671	0.000 214	0.005 348	6.46
B-SHOT	0.002 254	0.003 649	0.000 675	0.006 578	7.95
CI-SHOT	0.004 911	0.005 779	0.000 484	0.011 174	9.58
GRAY-SHOT	0.002 194	0.005 444	0.000 287	0.007 926	13.51
FPFH		0.038 144	0.000 406	0.038 550	46.60
SHOT		0.082 479	0.000 244	0.082 722	100.00

表4 特征描述子占用内存比较

Tab. 4 Comparison of memory consumption of different feature descriptors

描述子名称	内存占用/bit	内存占用比例/%
SD-SHOT	352	3.13
B-SHOT	352	3.13
CI-SHOT	704	6.25
GRAY-SHOT	704	6.25
FPFH	1056	9.38
ROPS	4320	38.40
SHOT	11264	100.00

3.4 实际机器人自动作业系统测试

在本节中将所提出的SD-SHOT应用于实际的机器人物体分拣系统中,验证自适应二进制简化描述子SD-SHOT的有效性。实验中,设置分别使用未进行简化的描述子SHOT、固定比率二进制简化描述子B-SHOT算法以及自适应简化描述

子SD-SHOT算法对实际采集的数据进行对比测试,验证二进制简化算法的效果以及意义。本文的实验平台和设计的物体分拣系统的系统结构分别如图3和图4所示。

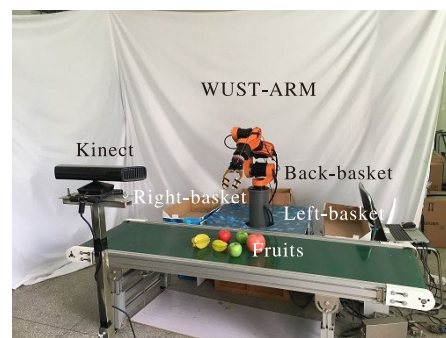


图3 物体分拣系统实验平台

Fig. 3 Experimental platform of object sorting system

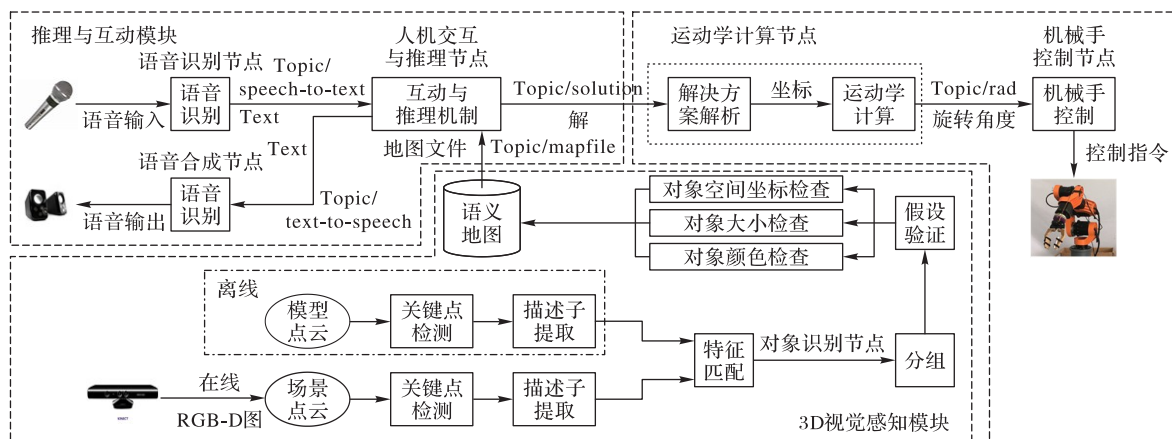


图4 物体分拣系统的系统架构

Fig. 4 System architecture of object sorting system

物体分拣系统主要有三个关键功能模块,分别是:三维视觉感知模块、推理与交互模块和自动编程与执行模块。三维视觉感知模块为整个系统提供高质量的环境语义信息;推理与交互模块将人机交互获得的规则与实时三维视觉感知接收到的实时语义图相结合,生成意图;自动编程与执行模块将具有语义信息的意图转换为可执行的机器人指令,控制机械手完成目标动作。在本项目的系统中,推理与交互模块是一种改进的CBR (Case-Based Reasoning)机制,在笔者之前发表的论文^[23]中将其称为“human-robot-environment”交互机制;而关于自动编程和执行模块,则在另一篇论文^[24]中做了详细的

介绍。

本小节的重点是验证SD-SHOT特征描述子在三维视觉感知模块中的有效性,因此本文假设用户是一次性正确地给出了一个完整的意图。评价指标为机器人能否根据三维视觉感知获得的信息实现正确的物体分拣,即正确识别物体的分类以及物体的位姿。

实验选取了5个物体模型来验证SHOT特征描述子和二进制简化特征描述子对物体的轮廓平滑性、对称性以及规则性欠佳的物体的识别效果。物体模型和场景由Kinect采集,具体物体分类如下:

物体类型 A: 轮廓对称物体——苹果。

物体类型 B: 轮廓存在起伏变化的对称物体——圆椒。

物体类型 C: 轮廓存在起伏变化的非对称物体——杨桃。

物体类型 D: 轮廓平滑的非对称物体——杯子(带把手)。

物体类型 E: 规则性欠佳的物体——香蕉

实际验证物体的 RGB 图和点云图如图 5 所示。



(a) 实际验证物体的 RGB 图



(b) 实际验证物体的点云图

图 5 实际实验图

Fig. 5 Images in actual experiment

针对所选择的每一种物体,实验通过人为随机改变物体位置和姿态采集 25 个模型场景对,采集 5 组数据,之后每组数据测试 10 次,合计测试 1 250 次。统计机器人分拣平台是否可以正确地识别并抓取物体,评价指标为分拣的平均正确率,结果如表 5 所示。从表 5 可以看出:对于二进制简化的特征描述子,本文提出的 SD-SHOT 在当前五个种类的物体中的识别正确率均优于 B-SHOT;而且相对于简化前的特征描述子 SHOT,SD-SHOT 在处理除了“轮廓存在起伏变化的对称物体”以外都具有相当或者略好的识别正确率。这是因为 SD-SHOT 采用了统计学中标准差的理论,在二进制量化时是依照数据的分布情况而自适应地二值简化,在这个过程中,对于特征明显的数(数据值大于均值与一个标准差的和)进行了放大使得数据的显著性得到了提升,从而提高了模型与场景之间物体的匹配正确率,最终提高了分拣平台分拣的正确率,验证了本文提出的简化方法的有效性。

表 5 实际作业平台物体分拣物体的平均正确率比较

Tab. 5 Average accuracy comparison of sorting objects on actual operating platform

描述子名称	平均分拣正确率/%				
	苹果	圆椒	杨桃	杯子	香蕉
SHOT	68	50	50	40	17
B-SHOT	60	31	50	60	25
SD-SHOT	72	44	50	60	42

4 结语

本文提出了一种基于标准差的自适应二进制简化方法,将它应用于 SHOT 特征描述子的简化中,并在 Bologna Kinect 数据集和实际机器人物体分拣平台对算法进行了评估。首先,通过改变本文提出的简化方法中的两个可变参数:简化单元 L 和标准差的个数 n 产生不同的简化特征描述子,并通过评价指标 T_{diff} 选择最优的参数组合。实验证明:当 $L = 352, n = 1$ 时,简化特征描述子的性能最好,将其命名为 SD-SHOT。此外,基于 PRC、关键点特征匹配的平均耗时以及特征描述子占用内存三个指标,将本文提出的 SD-SHOT 与 B-SHOT、GRAY-

SHOT、CI-SHOT 以及 SHOT 进行性能上的比较,结果表明,SD-SHOT 在标准数据集上综合性能优于 B-SHOT、GRAY-SHOT 和 CI-SHOT;而且,特征点匹配的平均耗时和特征描述子占用内存也最少(最优)。此外,本文还基于 SD-SHOT 构建了一个机器人物体分拣系统,实际系统验证表明采用本文方法简化后的特征描述子可以在实际场景中生成三维语义地图信息,能够满足实际实验验证阶段的作业需求。下一步会将本文的方法应用于更多的特征描述子的简化中,并且在其他数据集集中进行更广泛的测试;同时,将本文提出的方法推广于更多的实际应用场景。

参考文献(References)

- [1] LIU H, CONG Y, SUN G, et al. Robust 3-D object recognition via view-specific constraint [J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2020(Early Access): 1-11.
- [2] LI D, LIU N, GUO Y, et al. 3D object recognition and pose estimation for random bin-picking using Partition Viewpoint Feature Histograms [J]. Pattern Recognition Letters, 2019, 128: 148-154.
- [3] WANG F, LIANG C, RU C, et al. An improved point cloud descriptor for vision based robotic grasping system [J]. Sensors, 2019, 19(10): No. 2225.
- [4] RUSU R B, BLODOW N, MARTON Z C, et al. Aligning point cloud views using persistent feature histograms [C]// Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE, 2008: 3384-3391.
- [5] RUSU R B, BLODOW N, BEETZ M. Fast Point Feature Histograms (FPFH) for 3D registration [C]// Proceedings of the 2009 IEEE International Conference on Robotics and Automation. Piscataway: IEEE, 2009: 3212-3217.
- [6] TOMBARI F, SALTI S, DI STEFANO L. A combined texture-shape descriptor for enhanced 3D feature matching [C]// Proceedings of the 18th IEEE International Conference on Image Processing. Piscataway: IEEE, 2011: 809-812.
- [7] GUO Y, SOHEL F, BENNAMOUN M, et al. TriSI: a distinctive local surface descriptor for 3D modeling and object recognition [C]// Proceedings of the 2013 International Conference on Computer Graphics Theory and Applications/ International Conference on Information Visualization Theory and Applications. Setúbal: SciTePress, 2013: 86-93.
- [8] GUO Y, SOHEL F, BENNAMOUN M, et al. Rotational projection statistics for 3D local surface description and object recognition [J]. International Journal of Computer Vision, 2013, 105(1): 63-86.
- [9] KECHAGIAS-STAMATIS O, AOUF N. Histogram of distances for local surface description [C]// Proceedings of the 2016 IEEE International Conference on Robotics and Automation. Piscataway: IEEE, 2016: 2487-2493.
- [10] YANG J, CAO Z, ZHANG Q. A fast and robust local descriptor for 3D point cloud registration [J]. Information Sciences, 2016, 346/347: 163-179.
- [11] YANG J, ZHANG Q, CAO Z. Multi-attribute statistics histograms for accurate and robust pairwise registration of range images [J]. Neurocomputing, 2017, 251: 54-67.
- [12] YANG J, ZHANG Q, XIAO Y, et al. TOLDI: an effective and robust approach for 3D local shape description [J]. Pattern Recognition, 2017, 65: 175-187.

- [13] ZHAO B, LE X, XI J. A novel SDASS descriptor for fully encoding the information of a 3D local surface [J]. *Information Sciences*, 2019, 483: 363-382.
- [14] ZHAO B, XI J. Efficient and accurate 3D modeling based on a novel local feature descriptor [J]. *Information Sciences*, 2020, 512: 295-314.
- [15] GUO Y, BENNAMOUN M, SOHEL F, et al. A comprehensive performance evaluation of 3D local feature descriptors [J]. *International Journal of Computer Vision*, 2016, 116(1): 66-89.
- [16] 郝雯,王映辉,宁小娟,等. 面向点云的三维物体识别方法综述 [J]. *计算机科学*, 2017, 44(9): 11-16. (HAO W, WANG Y H, NING X J, et al. Survey of 3D object recognition for point clouds [J]. *Computer Science*, 2017, 44(9): 11-16.)
- [17] PRAKHYA S M, LIU B, LIN W. B-SHOT: a binary feature descriptor for fast and efficient keypoint matching on 3D point clouds [C]// *Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Piscataway: IEEE, 2015: 1929-1934.
- [18] PRAKHYA S M, LIU B, LIN W, et al. B-SHOT: a binary 3D feature descriptor for fast keypoint matching on 3D point clouds [J]. *Autonomous Robots*, 2017, 41(7): 1501-1520.
- [19] KECHAGIAS-STAMATIS O, AOUF N, CHERMAK L. B-HoD: a lightweight and fast binary descriptor for 3D object recognition and registration [C]// *Proceedings of the IEEE 14th International Conference on Networking, Sensing and Control*. Piscataway: IEEE, 2017: 37-42.
- [20] PRAKHYA S M, LIN W, CHANDRASEKHAR V, et al. Low bit-rate 3D feature descriptors for depth data from Kinect-style sensors [J]. *Signal Processing: Image Communication*, 2017, 51: 40-49.
- [21] DU L, MIN H, LIN Y. CI-SHOT: a statistical method for binary feature descriptor on 3D point clouds [C]// *Proceedings of the 2018 IEEE International Conference on Robotics and Biomimetics*. Piscataway: IEEE, 2018: 373-379.
- [22] LIN Y, SUN Y, MIN H. A general gray code quantized method of binary feature descriptors for fast and efficient keypoint matching [C]// *Proceedings of the 2nd International Conference on Intelligent Autonomous Systems*. Piscataway: IEEE, 2019: 1-7.
- [23] LIN Y, MIN H, ZHOU H, et al. A human-robot-environment interactive reasoning mechanism for object sorting robot [J]. *IEEE Transactions on Cognitive and Developmental Systems*, 2018, 10(3): 611-623.
- [24] LIN Y, ZHOU H, CHEN M, et al. Automatic sorting system for industrial robot with 3D visual perception and natural language interaction [J]. *Measurement and Control*, 2019, 52(1/2): 100-115.
- [25] TOMBARI F, SALTI S, DI STEFANO L. Performance evaluation of 3D keypoint detectors [J]. *International Journal of Computer Vision*, 2013, 102(1/2/3): 198-220.

This work is partially supported by the Hubei Natural Science Foundation Youth Project (2020CFB116), the Scientific Research Program for Young and Middle-Aged Talents of Department of Education of Hubei Province (Q20191108).

LIU Shuangyuan, born in 1995, M. S. candidate. His research interests include robot perception, three-dimensional vision.

ZHENG Wangli, born in 1985, M. S., engineer. His research interest includes test for power communication equipment.

LIN Yunhan, born in 1985, Ph. D., lecturer. His research interests include robot perception, robotic reasoning and control.