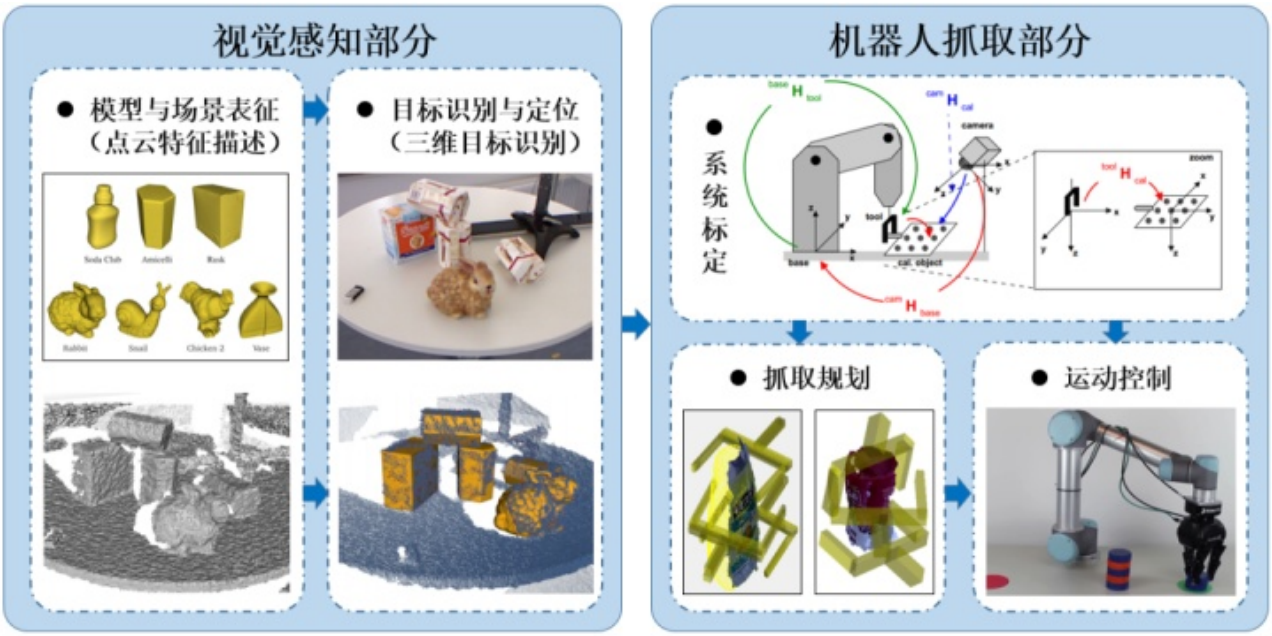
基于点云的机器人抓取识别综述

zhuanlan.zhihu.com/p/159467103



机器人作为面向未来的智能制造重点技术，其具有可控性强、灵活性高以及配置柔性等优势，被广泛的应用于零件加工、协同搬运、物体抓取与部件装配等领域，如图1-1所示。然而，传统机器人系统大多都是在结构化环境中，通过离线编程的方式进行单一重复作业，已经无法满足人们在生产与生活中日益提升的智能化需求。随着计算机技术与传感器技术的不断发展，我们期望构建出拥有更加灵敏的感知系统与更加智慧的决策能力的智能化机器人系统。

图1-1 机器人的应用领域

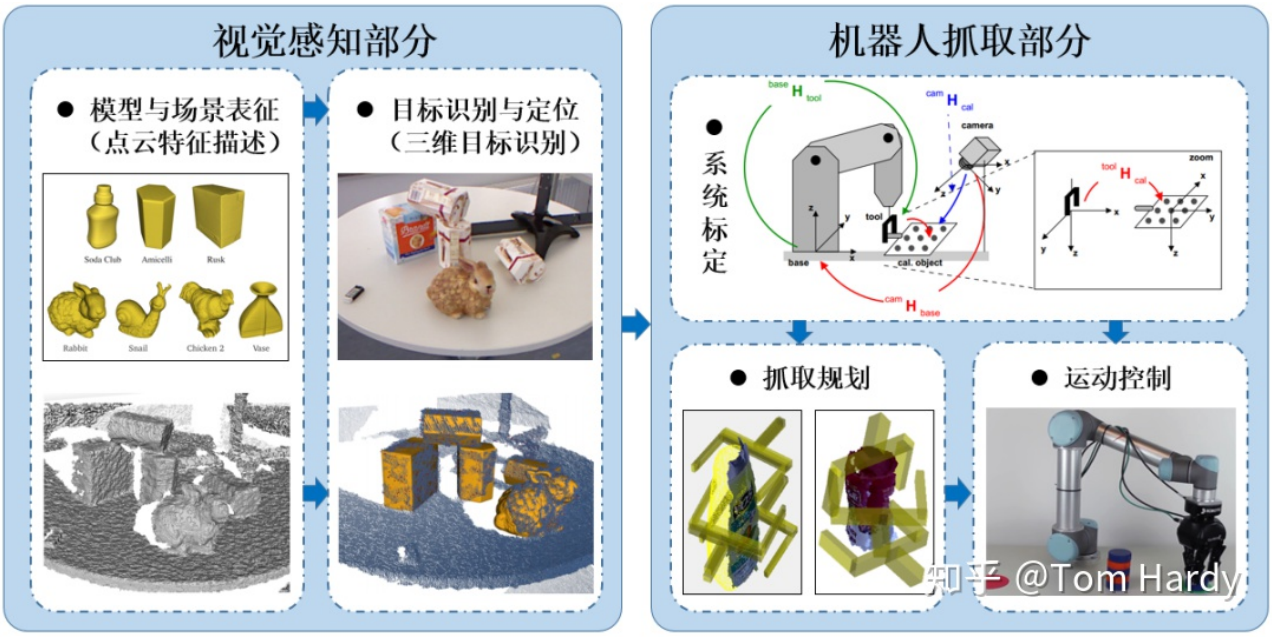
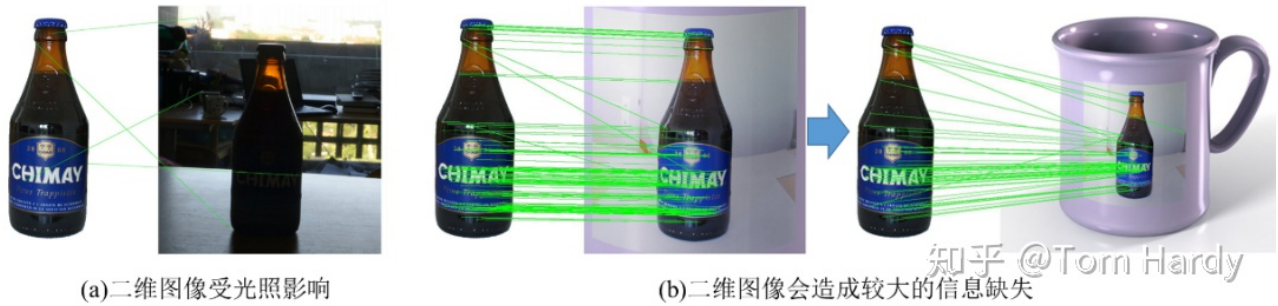


图1-2 机器人抓取的操作流程与步骤

机器人抓取与放置是智能化机器人系统的集中体现，也是生产与生活中十分重要的环节，近几年来在工业界与学术界得到了深入而广泛的研究。具体的机器人抓取可以分为视觉感知部分与机器人抓取操作部分。视觉感知部分又包含：模型与场景表征、目标识别与定位这两个步骤；而机器人抓取操作部分则包含：系统标定、运动控制与抓取规划等步骤，如图1-2所示。这其中，机器人通过视觉传感器感知环境并实现对目标物体的识别与定位，也就是视觉感知部分，是十分重要的环节，其直接决定了后续机器人的抓取精度。

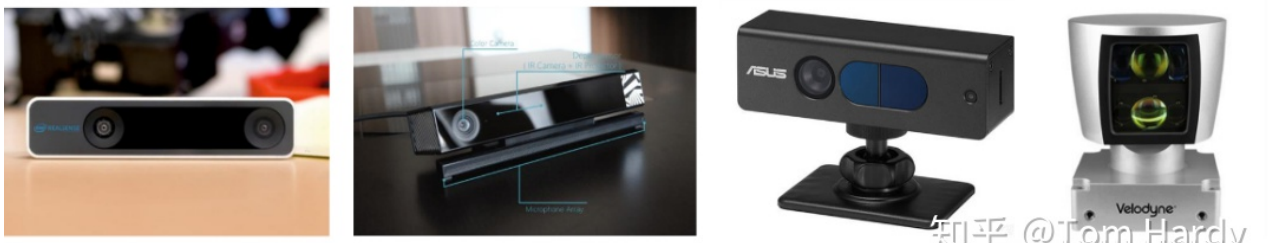


(a)二维图像受光照影响　　(b)二维图像会造成较大的信息缺失

图1-3 二维图像的部分缺陷

受益于计算机算力的不断提高以及传感器成像水平的高速发展，目前针对结构化环境或者半结构化环境中，基于二维图像的机器人平面单目标物体的抓取技术已经趋于成熟，并取得了丰富的研究成果[1][2][3]。然而，对于现实复杂环境中的三维物体，仅使用二维信息对三维目标进行表征，会不可避免的造成信息损失，如图1-3所示，从而难以实现非结构化环境中机器人对于多目标物体的高精度抓取操作。因此，如何提升机器人的视觉感知能力，并基于此在复杂环境中自主完成对目标物体的识别、定位、抓取等操作是一个很有价值的研究问题。

近年来，随着低成本深度传感器（如Intel RealSense、Xtion以及Microsoft Kinect等）与激光雷达的飞速发展，如图1-4所示，三维点云的获取越来越方便。这里的点云实际上就是在相机坐标系下，对所拍摄的物体或者场景表面进行点采样。物体对应的点云数据在在数学上可以简单的理解为三维坐标的无序集合。三维点云数据相对于平面二维图像具有如下优势：（1）可以更加真实准确的表达物体的几何形状信息与空间位置姿态；（2）受光照强度变化、成像距离以及视点变化的影响较小；（3）不存在二维图像中的投影变换等问题。三维点云数据具有的以上优势使得其有望克服平面二维图像在机器人目标识别与抓取中存在的诸多不足，所以其具有很重要的研究意义以及广泛的应用前景。因此，近年来针对点云的视觉研究以及基于点云的机器人抓取成为了机器人领域新的研究热点。



(a) Inter RealSense　(b) Microsoft Kinect V2　(c) Xtion　(d) Velodyne 激光雷达

图1-4 点云获取设备示意图

对应前文的，在基于点云的机器人抓取可以分为点云特征描述（模型与场景表征）、三维目标识别（目标识别与定位）与机器人抓取操作这三个部分[39][40]。进一步的，点云特征描述指的是，将模型与场景对应的无序点集通过特定的算法编码为低维的特征向量，用此来表征对象的局部或者全局信息，其应当具有足够的描述力与稳定性。三维目标识别则主要是指，利用模型与场景的表征结果，在场景中识别出目标物体，并估计出其对应的位置与姿态。对于特征描述与目标识别，尽管现有文献提出了不少算法，并且在特定的环境中取得了不错的效果，然而如何在包含噪声、干扰、遮挡与密度变化的复杂非结构化环境中提取有效而稳定的特征，实现对多目标物体的准确识别定位以及高精度抓取，仍然是极富挑战性的一个问题[4]。

综上所述，基于点云的机器人抓取作为智能化机器人系统的集中体现，近几年来得到了工业界和学术界的广泛关注，并围绕点云特征描述、三维目标识别与机器人抓取操作这三个方面展开了深入研究。具体的，在点云特征描述部分，主要关注描述子的鉴别力、鲁棒性、计算效率与紧凑性等性能；在三维目标识别部分，主要关注目标的识别准确率与定位精度问题；而在机器人抓取操作部分，抓取系统的参数标定与多目标物体的数据分析都是很重要的环节。

1.1国内外研究现状

受益于点云数据自身的优势、计算机算力的不断提高与传感技术的不断发展，基于点云的机器人抓取成为了机器人领域新的研究热点，具有十分诱人的研究价值与应用前景。近年来，学术界与工业界围绕基于点云的机器人抓取，在点云特征描述、三维目标识别与机器人抓取操作这三个方面展开了广泛而深入的研究，取得了显著进展，下面分别从上述三个方面进行文献综述。

1.1.1点云特征描述

点云特征描述在机器人抓取中主要是应用于视觉感知部分的模型与场景表征。一种合格的特征描述算法应该有较高的描述力来表征对应的局部点云表面。此外，此外其还应该对于点云噪声、表面孔洞、部分遮挡、视点改变以及分辨率变化等稳健[4]，如图1-5所示。

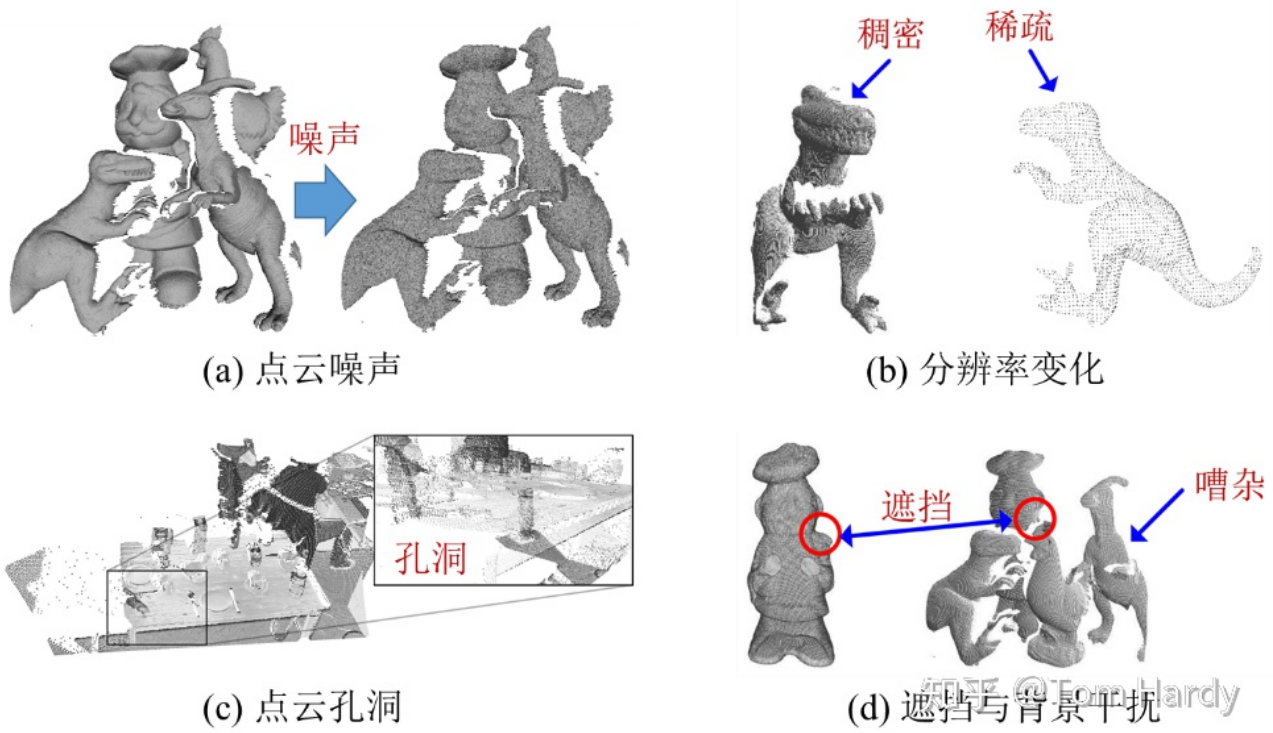


图1-5 点云场景存在的挑战

现有的特征描述算法可以分为全局特征和局部特征两大类[5]。全局特征采用模型的整体几何信息构建得到，典型代表有Osada等[6]提出来的Shape distribution描述子，Wahl等[7]提出来的SPR（Surflet-pair-relation）描述子以及Funkhouser等[8]提出来的Spherical harmonics描述子。全局描述子拥有较高的计算效率和分类能力，但是其对于遮挡比较敏感，很难用于目标识别和精确定位[9]。鉴于此，局部点云的概念被提出，局部特征描述算法得到了深入的研究和广泛的关注。其首先提取关键点建立局部邻域，根据邻域内各点的空间分布信息和几何特征构建描述矩阵。局部描述子对于背景干扰和遮挡鲁棒，相比于全局描述子更适合用于非结构化环境中的目标识别[4][10]。

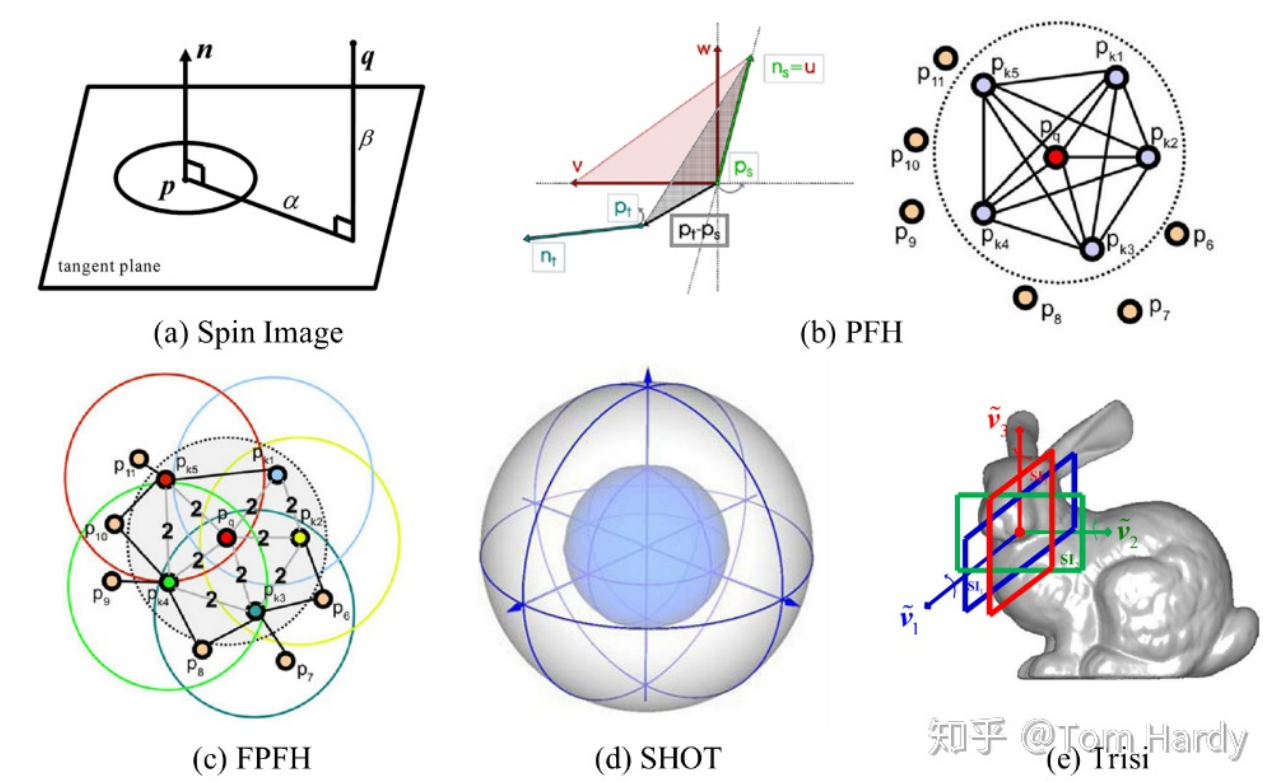


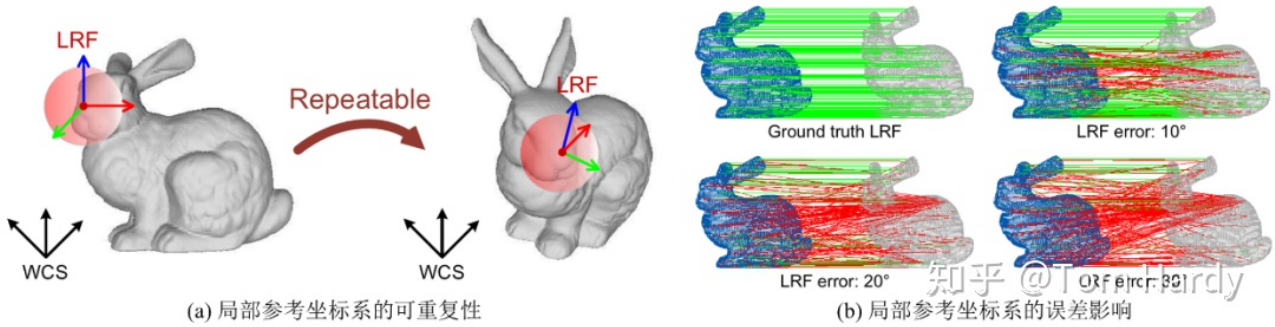
图1-6 部分局部描述算法示意图

局部描述算法又可以根据有无建立局部参考坐标系(Local Reference Frame, LRF)进行分类[11]。不依赖LRF的特征描述子都是使用局部几何信息的统计直方图或者信息量来构成特征矩阵[12]。例如，Johnson等[13]提出了Spin image描述算法，如图1-6(a)，它首先以关键点的法线作为参考轴，用两个参数对关键点的每个邻域点进行编码，然后根据这两个参数将局部邻域点进行分箱，进而生成一个二维直方图。Spin image描述子已经成为了三维特征描述子评估体系的实验基准[4][14]。但是，其存在诸如对数据分辨率变化和非均匀采样敏感等缺陷[15]。Rusu等[16]提出了PFH（Point Feature Histogram）描述算法：其对于关键点邻域内的每一个点对，首先建立Darboux框架，然后采用[7]中的方法计算由法向量和距离向量得到的四个测量值，最后将所有点对的测量值进行累加生成一个长度为16的直方图，如图1-6(b)。为了降低计算复杂度，Rusu[17]等仅将关键点与其邻域点之间的测量值进行累加，随后进行加权求和得到FPFH（Fast-PFH），如图1-6(c)。FPFH保留了PFH的绝大部分鉴别信息，但是其对于噪声敏感[5]。目前绝大多不依赖于LRF的描述子仅利用了点云的部分几何特征，而很难编码局部空间分布信息，因而其都鉴别力不强或者鲁棒性较弱[15]。

对于建立了局部参考坐标系的描述子，则利用定义的LRF来同时对空间分布信息和几何特征进行编码以提高其鉴别力和鲁棒性[18]。例如，Tombari等人[19]首先利用加权主成分分析（PCA）的方法为关键点构建了一个局部参考坐标系，进而在该LRF下将关键点对应的球形R-近邻空间进行栅格化处理，然后依据关键点法线与落入每一个子单元的点法线间的夹角将这些点累积到一个数据统计直方图中，最后串联各个直方图便获得SHOT（Signatures of Histograms of Orientation）特征，如图1-6(d)。SHOT计算效率高，但是对于分辨率变化敏感[5]。Guo等[18]通过计算局部表面对应散布矩阵的特征向量来建立LRF，然后利用旋转投影的方法对三维点集进行降维并建立分布矩阵，之后提取分布矩阵的信息量生成最后的RoPS（Rotational Projection Statistics）描述子。RoPS有着优越的综合性能[5]，但是其只能用于mesh网格文件，也就是说其无法作用于原始的

xyz点云数据[20]。并且，其将数据投影到了二维平面会造成较大的信息损失[21]。之后，Guo[15]在RoPS的LRF算法基础上进行改进，提高了稳定性，然后在坐标系的每一个参考坐标轴上求取局部邻域的Spin Image特征，串联组成Trisi（Triple-Spin Image）局部特征描述子，如图1-6(e)。基于LRF的局部描述算法的鉴别力和鲁棒性很依赖于所建立的局部参考坐标系的可重复性与稳定性，如果坐标系存在轻微的偏差，会对最终的描述向量造成严重的影响[22]，如图1-7。然而，目前已有的局部坐标系算法存在可重复性差或者方向歧义的问题[23]。

综上所述，对于不建立局部参考坐标系的特征描述子，由于不能融入空间分布信息，普遍存在鉴别力不高、对于噪声比较敏感等问题；而拥有局部参考坐标系的特征描述子的描述力和鲁棒性则主要依赖于所对应的坐标系建立算法，然而目前已有的坐标系建立方法均存在可重复性差或者方向歧义的问题[22]，相应的特征提取算法在鉴别力、鲁棒性与计算效率方面依然有提升的可能[5]。



(a) 局部参考坐标系的可重复性　　(b) 局部参考坐标系的误差影响

图1-7 LRF的误差影响

1.1.2三维目标识别

在基于点云的机器人抓取领域，完成了模型与场景的表征，下一步则是进行目标识别与定位，也就是在点云场景中对待抓取模型进行三维目标识别以及对应的姿态估计。现有的三维目标识别算法主要包括基于局部特征的算法、基于投票的算法、基于模板匹配的算法以及基于学习的方法[24][25]。

基于局部特征的目标识别算法则主要分成五个部分：关键点检测、特征提取、特征匹配、假设生成、假设检验[26][27]。在这里关键点检测与特征提取组合对应的就是进行模型与场景表征。由于点云的点集数量巨大，如果对每个点都进行特征提取则会造成计算机算力不足的情况，因此会在原点云中提取稀疏而区分度高的点集作为关键点。关键点应当满足可重复性和独特性这两个重要属性[28]。前者涉及的是在各种干扰下（噪声、分辨率变化、遮挡与背景干扰等）可以精确提取相同关键点的能力；而后者则是指提取的关键点应当易于描述、匹配与分类[29]。在点云领域，经典的关键点提取算法包括Harries 3D[30]，ISS（Intrinsic Shape Signature）算法[31]，NARF（Normal Aligned Radial Feature）算法[32]。特征提取部分则主要是在物体表面提取稳固的局部特征，详见本章1.3.1部分的讨论。

特征匹配的作用则是建立一系列的关键点特征对应关系，如图1-8所示。经典的特征匹配算法有最近邻距离比值（NNDR）、阈值法、最近邻策略（NN）等[33]。论文[33]则表明NNDR与NN的匹配算法优于阈值法的匹配效果，NNDR亦是目前使用最多的匹配策略

[34]。为了降低计算复杂度，一般都会使用高效的搜索算法来优化特征匹配，使其快速地找到场景特征库中与当前特征对应的k近邻特征。常用搜索算法包括k-d树[35]、局部敏感树[31]、哈希表[36]与二维索引表[37]等。

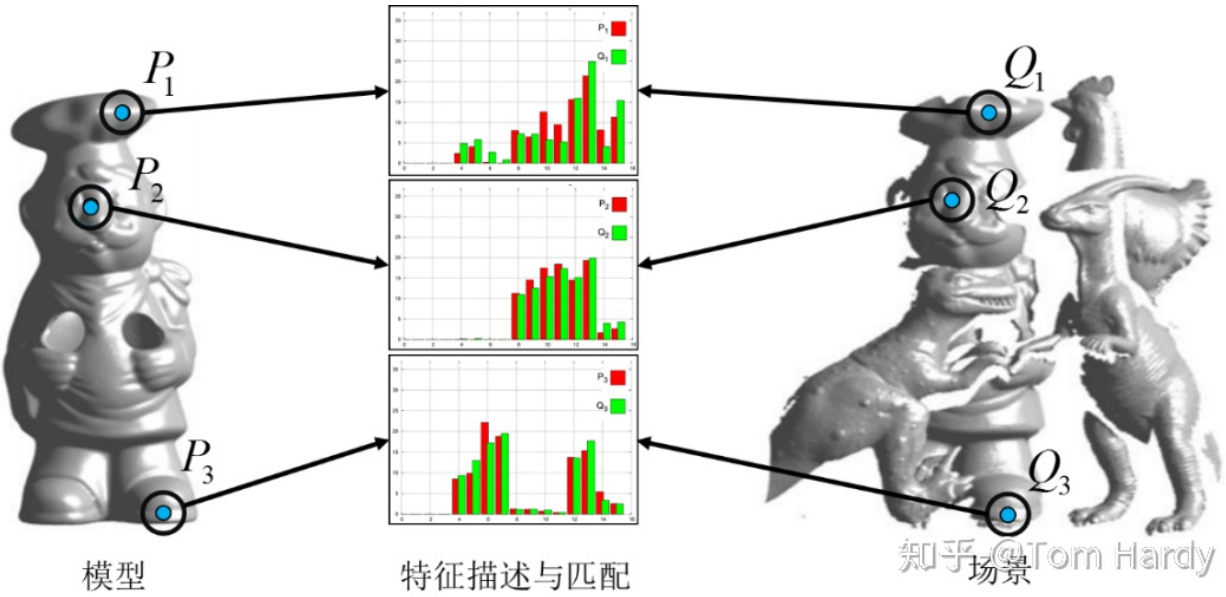


图1-8 局部特征匹配过程示意图

假设生成部分则主要是利用匹配上的特征对集合找出在场景中可能的模型位置，并建立对应的姿态估计（即计算变换假设）[38]。值得注意的是，在匹配上的特征对集合中，既会存在正确的特征对，也会有大量有误差的特征对。因此在计算变换假设的时候，需要使用有效的算法策略尽可能的剔除错误特征对，从而得到较为准确的模型与场景间的变换关系。这一部分的方法主要包括随机一致性采样（RANSAC）、姿态聚类、几何一致性以及扩展霍夫变换等。RANSAC算法首先随机选取k组特征对来计算模型到场景间的变换矩阵（这里k为生成一个变换矩阵所需要的最少特征对数量），并统计满足这个变换矩阵的点对数量。使用这个算法的论文包括[38][39][40]。姿态聚类算法则认为当模型在场景中被正确识别后，大多数模型与场景对齐的假设生成变换矩阵都应当在真实的位姿矩阵（ground truth）附近。使用这个算法的论文包括[31][41][42]。几何一致性技术则认为如果特征对不满足几何约束关系则会使得估计出来的变换矩阵有较大的误差，所以希望使用几何约束来剔除误差较大的匹配点对，进而提高生成的变换矩阵的准确性。使用该算法的论文包括[13][43][44]。扩展霍夫变换则是利用特征对间的平移和旋转等参数构成广义的霍夫空间，然后进行投票统计。这个广义的参数化霍夫空间中的每一个点都对应模型与场景间的一组变换关系，空间中的峰值点被认为是模型到场景变换矩阵估计的最优解。采用这种算法的论文包括[45][46][47]。

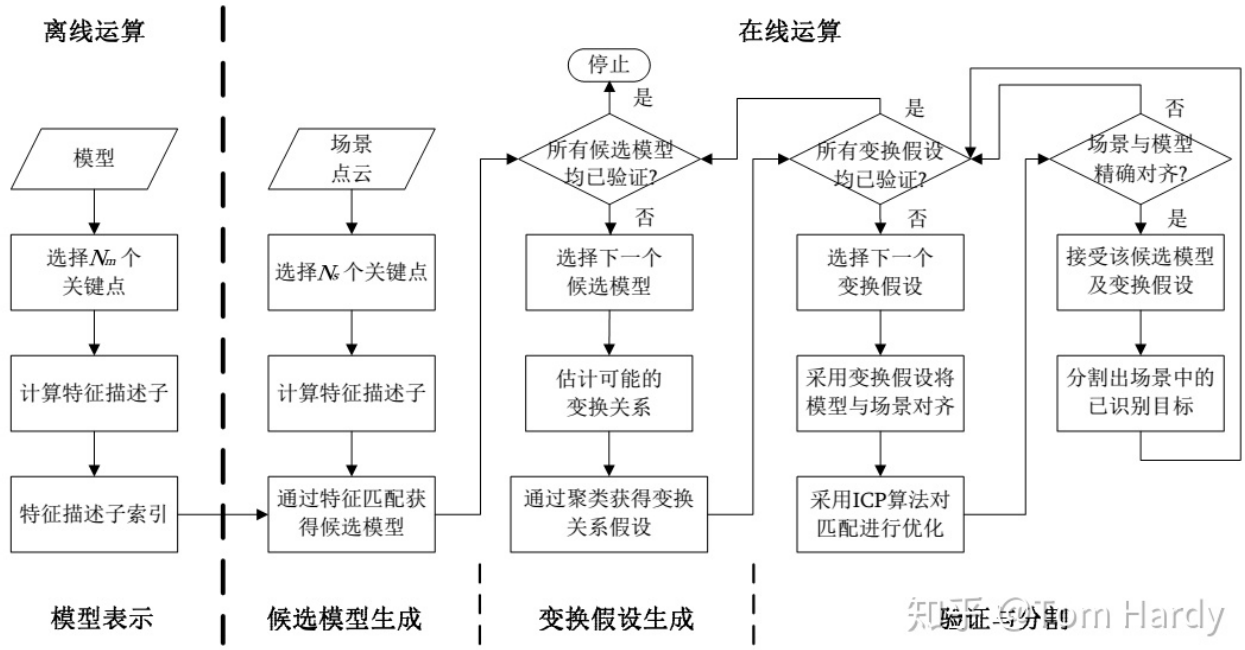


图1-9 基于特征提取的目标识别流程图

假设检验部分则是为了得到假设生成部分所计算出来的潜在变换关系中真正正确的变换矩阵。Hebert与Johnson[13][48]采用模型与场景的对应点数和模型总点数的比值作为相似度参数。当相似度大于设定的阈值时，则认为当前的变换矩阵是正确的。Main[49]则采用特征相似度与点云匹配精度作为综合评价指标。Bariya[43]首先计算出模型与场景的交叠面积，并将模型可见面积和重叠面积的比值作为相似度度量。Papazov[40]则提出了一个包含惩罚项和支持项的接收函数用于评估建设生成的姿态变换质量。Aldoma[44][26]则建立了场景到模型的拟合、模型到场景的拟合、遮挡关系以及不同模型间的关联这几个条件建立了一个代价函数，然后通过求取这个函数的极小值来获得理论上最优的变换姿态。

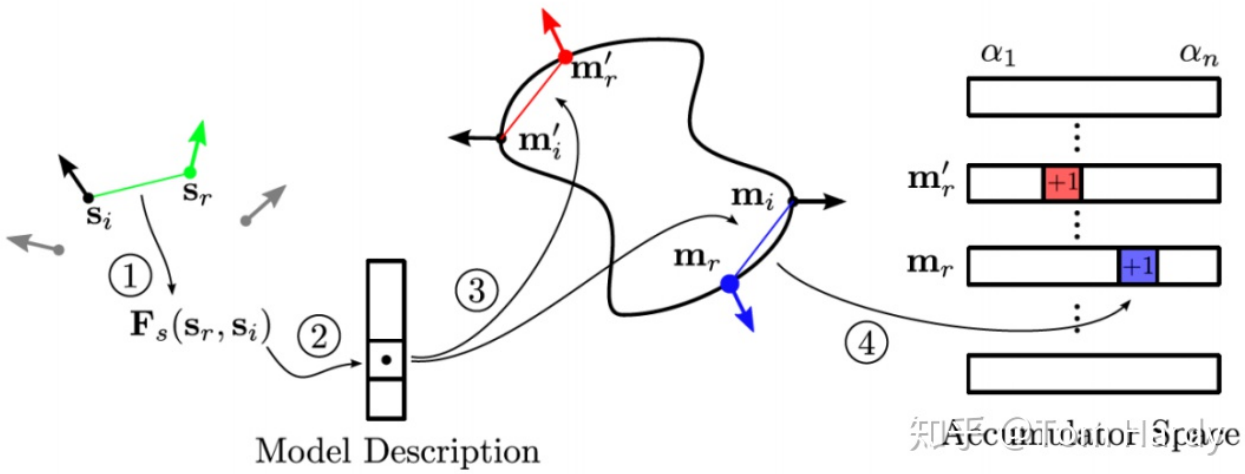


图1-10 PPF投票算法示意图

基于投票的三维目标识别算法则是直接匹配模型与场景间的固有特性，生成有限的候选姿态集后，利用先验条件构造支持函数与罚函数并对每一个姿态进行投票，进而得出最优的变换矩阵。Drost等人[41]提出了用于目标识别的点对特征（Point Pair Features, PPF），这也是三维目标识别领域的经典算法，算法原理如图1-10所示。其利用了点对间

最为朴素的特征：距离与法线夹角，构造出有四个参数的特征数组；然后结合哈希表进行穷举匹配，利用高效的投票方案得出最优的姿态估计。Kim等人[50]则在原始PPF特征中加入了可见性特征（空间、表面与不可见表面），增强了PPF的匹配能力。Choi等人[51]在此基础上提出了对点对特征进行分类的策略，如边界上的点对或者是由边缘点组成的点对等。利用这种分类方法可以减少训练和匹配的特征数量，加快了匹配速度以及投票效率。此外，Choi等人[52]还在PPF的点对特征上加入了颜色分量，创建了Color-PPF，实验结果表明其识别率明显提高。随后，Drost等人[53]又提出了利用几何边缘（边界和轮廓）来计算PPF，这种算法显著改进了在高度遮挡场景中的识别率。Birdal等人[54]则提出了先对场景进行分割，在进行PPF匹配的识别策略。更进一步的，Hinterstoisser等人[55]针对PPF提出了一种新的采样方法以及一种新的姿态投票方案，使得这种算法对噪声和背景干扰更加稳健。Tejan等人[56]则从RGB-D图像中训练了一个霍夫森林，在树中的叶子上存储着目标识别6D姿态的投票。

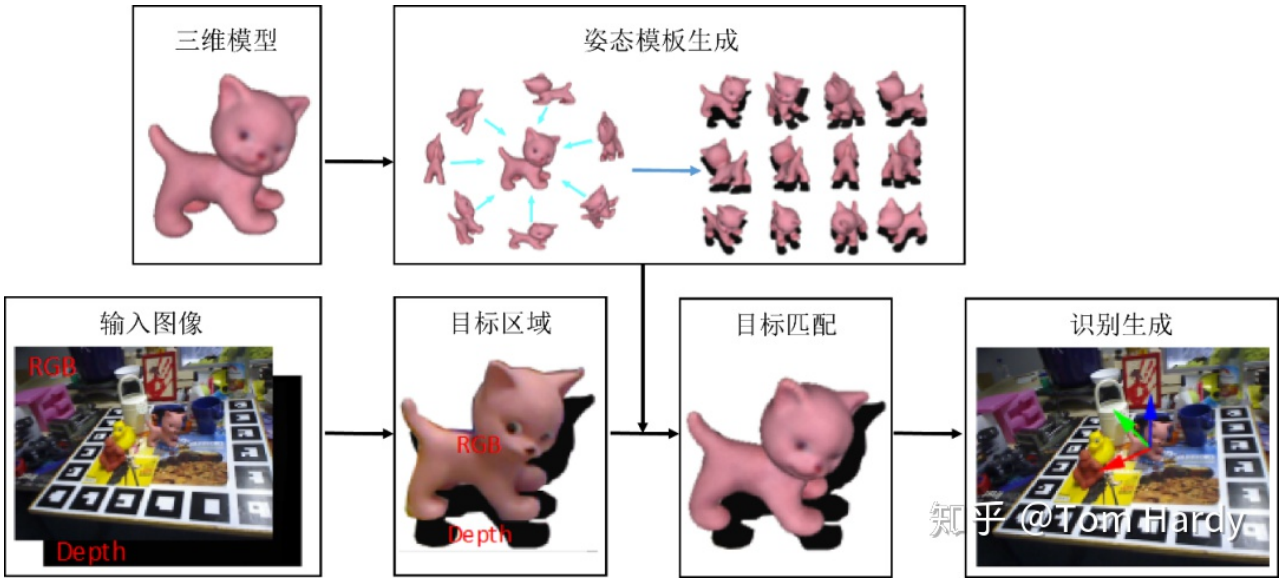


图1-11 基于模板匹配的目标识别流程

基于模板匹配的目标识别算法则主要是针对无纹理物体的检测。其利用已有的三维模型从不同的角度进行投影，生成二维RGB-D图像后再生成模板；然后将所有的模板与场景匹配，进而得出最优的模型位姿，算法原理如图1-11。Hinterstoisser等人[57]提出了经典的Linemod算法，其结合了彩色图像中的梯度信息再结合深度图像中的表面法线信息生成图像模板，在场景图像中利用滑窗搜索的方式进行模板匹配。Hodan等人[58]提出了一种实用的无纹理目标检测方法，也是实用滑动窗口的模式，对于每个窗口进行有效的级联评估。首先通过简单的预处理过滤掉大部分位置；然后对于每一个位置，一组候选模板(即经过训练的对象视图)通过哈希投票进行识别；最后通过匹配不同模式下的特征点来验证候选模板进而生成目标的三维位姿。

基于学习的方法，Brachmann等人[59]提出的基于学习的目标识别算法，对于输入图像的每一个像素，利用其提出的回归森林预测待识别对象的身份和其在对象模型坐标系中的位置，建立所谓的“对象坐标”。采用基于随机一致性采样算法的优化模式对三元对应点对集进行采样，以此创建一个位姿假设池。选择使得预测一致性最大化的假设位姿作为最终的位姿估计结果。这个学习模型在论文[60]中得到了多种扩展。首先，利用auto-context算法对于随机森林进行改进，支持只是用RGB信息的位姿估计；其次，该模型不仅考虑已知对象的位姿，同时还考虑了没有先验模型库的目标识别；更多的，其使用随

机森林预测每一个像素坐标在目标坐标系上的完整三维分布，捕捉不确定性信息。自从深度卷积神经网络（DCBB）[61]提出以来，基于深度学习的方法近年来变得十分流行，例如RCNN[62]，Mask-RCNN[63]，YOLO[64]与SSD[65]等。最近的综述论文[66]对于这些算法进行了详细的阐述和比较。

综上所述，在目前已有的目标识别算法中，基于几何一致性与随机一致性采样的管道方法存在组合爆炸的问题，其对应的计算复杂度为O(n3)；而基于点对特征的目标识别方案则会由于法线方向的二义性问题造成识别的准确率下降，并且其对应的计算复杂度为O(n2)；基于模板匹配的目标识别算法(Linemod)则存在对于遮挡敏感等问题。虽然各种算法在特定的数据集上都取得了不错的效果，但是在非结构化环境中的目标识别准确率依然有较大的提升空间。

1.1.3机器人抓取操作

1 条评论