

室内服务机器人实时目标识别与定位系统设计

黄海卫¹, 孔令成^{2,3}, 谭治英^{2,3}

(1. 常州大学 信息科学与工程学院, 江苏 常州 213164; 2. 中国科学院 合肥物质科学研究院, 安徽 合肥 230031; 3. 常州先进制造研究所, 江苏 常州 213164)

摘要: 针对服务机器人在室内复杂环境下难以准确获得物体标识和空间信息的问题, 提出一种结合三维点云分割和三维全局特征匹配的实时物体识别与定位系统。对获取的点云数据进行快速有效分割, 计算场景内最大水平面, 通过改进的聚类算法提取水平面上的物体聚类, 采用基于聚类视点特征直方图 (CVFH) 特征匹配的方法识别目标物体, 计算目标所在空间位置。实验结果表明, 该识别定位系统满足服务机器人自主实时准确抓取的要求。

关键词: 服务机器人; 物体识别; 分割; 三维全局特征; 聚类

中图分类号: TP242.6⁺2 文献标识码: A 文章编号: 1000-7024 (2016) 08-2228-05

doi: 10.16208/j.issn1000-7024.2016.08.046

Real-time object recognition and localization system for indoor service robot

HUANG Hai-wei¹, KONG Ling-cheng^{2,3}, TAN Zhi-ying^{2,3}

(1. College of Information Science and Engineering, Changzhou University, Changzhou 213164, China;
2. Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China;
3. Institute of Advanced Manufacturing Technology of Changzhou, Changzhou 213164, China)

Abstract: Aiming at the problem that service robots in complex environment are difficult to accurately obtain object identification and spatial information, the real-time object recognition and localization system was proposed which integrating 3D point clouds segmentation and global feature matching. Segmentation was carried out effectively and efficiently to find the maximum horizontal surface of the primary point cloud data. The improved clustering algorithm was used to extract the objects, and objects were identified using the CVFH feature matching based approach. Experimental results show the system can well meet the real-time and accuracy requirements.

Key words: service robot; object recognition; segmentation; three-dimensional global feature; cluster

0 引言

服务机器人的基本功能之一是搬运物体, 而一整套自主搬运动作的完成离不开对目标的准确识别与实时定位, 因此对识别定位方法的研究具有重要现实意义, 也是服务机器人领域的重要课题之一。

为了能够让手臂精准的抓取目标, 识别系统不仅要提供目标的标识信息, 还要提供目标的空间信息。由于室内机器人所处的环境相对复杂且多变, 在这样的环境下实现实时识别和定位物体是一个严峻的任务。文献 [1] 利用改

进的颜色直方图方法来识别物体, 图像处理速度比较快, 但光照变化对其影响明显^[1]; 文献 [2] 提出用 SURF 特征匹配, SURF 是 SIFT 的进化版, 与 SIFT 相比具有更快的计算速度, 在二维图像识别领域具有很好的应用, 但一般通过图像匹配难以获取准确的物体位置信息^[2]; 文献 [3] 提出用 NARF (径向归一化描述) 描述物体来进行识别, 但计算复杂难以满足实时性^[3,4]。

本文设计的系统采用 Kinect 实时采集室内环境点云数据。对数据进行预处理去除部分噪声点, 并分割计算最大水平面, 提取水平面上的物体聚类, 计算各个聚类的

收稿日期: 2015-08-16; 修订日期: 2015-10-26

基金项目: 青年科学基金项目 (61401437); 中科院合肥研究院院长基金项目 (YZJJ201322); 江苏省科技支撑计划基金项目 (SBE201201023)

作者简介: 黄海卫 (1989-), 男, 江苏海门人, 硕士研究生, 研究方向为机器视觉; 孔令成 (1965-), 男, 安徽合肥人, 硕士, 研究员, 研究方向为机器人技术、智能传感器; 谭治英 (1985-), 女, 山东安丘人, 博士, 助理研究员, 研究方向为图像处理、机器视觉。

E-mail: lckong@iim.ac.cn

CVFH 特征^[5,6], 进行特征匹配识别目标, 由于点云数据包含目标的空间信息, 可以将目标的中心点坐标作为目标在相机坐标系下的位置。实验结果表明, 该系统识别率高, 返回的位置信息准确, 同时具有较高的实时性。

1 系统结构

本文所设计的系统分为两部分如图 1 所示, 第一部分

离线模板库建立, 对所识别的目标建立特征模板库。第二部分为在线识别, 首先对获取的场景点云进行预处理, 通常室内机器人所要抓取的物体都在类似桌面的平面上, 因此我们可以先将点云中的最大水平面找出来, 然后将平面上的点进行聚类得到相互独立的物体。提取每个聚类的 CVFH 特征并和模板库中的特征进行匹配^[7-9], 标出识别物体, 最后计算物体在摄像机坐标系的空间坐标。

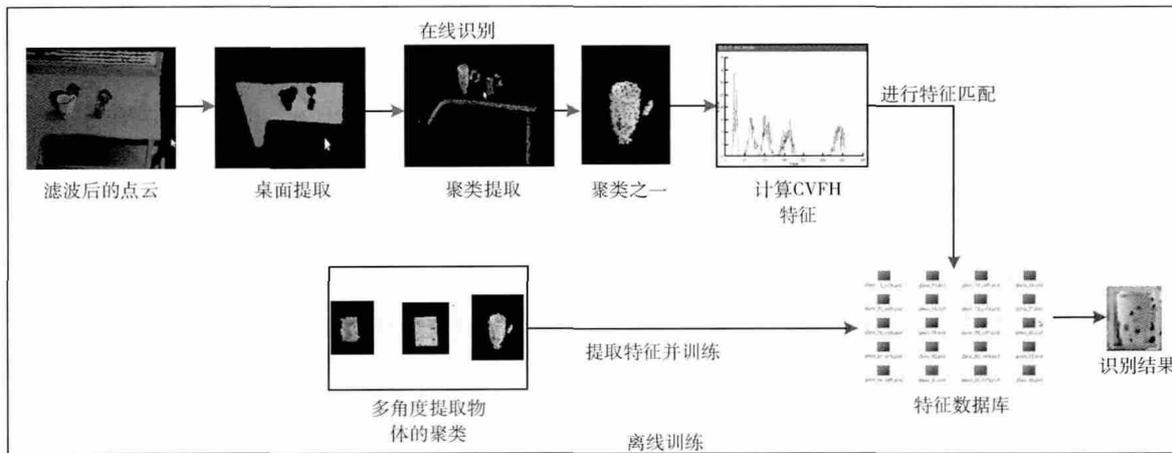


图 1 系统总体框架

2 系统实现

机器人所处的室内环境虽然复杂, 但并非看到的场景都需要去处理, 这里我们只关注手臂工作的区域, 也就是类似于桌面的平面上, 因此在提取特征之前, 要进行有效的分割。

2.1 平面提取

通常从三维点云中提取水平面都是采用基于随机采样一致性 (RANSAC) 的方法^[10], 即每次随机选取 3 个点, 计算由这 3 个点构成的平面模型, 然后按照给定的阈值去拟合其它点, 采用该方法提取的平面带有较多噪声点, 且时间复杂度较高。本文先利用主成分分析法 (PCA) 估计局部邻域点云的法向量, 然后计算该邻域的粗平面^[11]。

对于一个包含 N 个点的局部邻域, S 为其近似平面, μ 为三维质心, 平面法向量 \bar{n} 可通过 PCA 方法的协方差矩阵算得, 计算公式为

$$E = \frac{1}{k} \sum_{i=1}^k (N_i - \mu) * (N_i - \mu)^T \quad (1)$$

对协方差矩阵进行特征值分解, 得到特征值、特征向量和三维中心 μ , 则近似平面 S 可表示为

$$Ax + By + Cz + D = 0 \quad (2)$$

其中, D 值, 近似得到平面 S 。对原始点云 P_d 进行上述粗提取得到平面集合 S_i , 由于抓取的物体在水平面上, 所以取法向量平行于世界坐标系 Z 轴的平面即 $\bar{n}_i \times \bar{z} \approx 0$ 的平面, 然后采用平面区域增长法合并位于近似平面上的点。

对于任意一点 $P_i(x_i, y_i, z_i)$ 到平面 S_i 的距离 d_s 和到三维质心 μ_i 的距离 d_μ , 两者计算公式分别为

$$d_s = \frac{ax_i + by_i + cz_i + d}{\sqrt{a^2 + b^2 + c^2}} \quad (3)$$

$$d_\mu = \sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2} \quad (4)$$

通过和给定的阈值进行比较来判断是否属于平面, 以及属于哪个平面, 最后比较所有平面聚类的大小, 平面聚类最大者即为所要寻找的水平面^[12], 算法的具体步骤如下:

- (1) 选取局部点集 M , 构建协方差矩阵 C , 计算得到的最小特征值以及对应的最小特征向量, 然后拟合初始平面集 S_i ;
- (2) 初始化一个空的平面集合 T_i , 计算平面集 S_i 与 Z 轴的关系, 如果 $\bar{n}_i \times \bar{z} \approx 0$, 将 S_i 加入 T_i 中, 否则剔除;
- (3) 进行平面区域增长, 遍历面集以外的点云, 计算当前点 P_0 到平面 T_i 的距离 d_T 以及到局部邻域质心的距离 d_μ , 如果 $d_T < d_{T(thresh)}$ 并且 $d_\mu < d_{\mu(thresh)}$ 则将点 P_0 加入到对应平面中, 重复步骤 (3) 直到没有点满足为止;
- (4) 步骤 (3) 区域增长结束, 判断平面集 T_i 的大小, 得到最大平面 T_{max} 。

按照上面的算法步骤得到结果如图 2, 图 3 所示。

2.2 物体聚类

分割最大水平面 T_{max} 上的点云前, 先确定剩余点集中属于水平面的点, 利用点云凸包算法计算水平面 T_{max} 的二维凸包, 然后将剩余点云沿 Z 轴向下投影以确定水平面上点云 Q 的大小。



图 2 预处理后的点云



图 3 本算法找到的平面



图 5 分割不足结果



图 6 本文算法结果

欧式聚类是最常用的物体分割方法之一^[13,14]，该方法将空间一定距离内的点合并为一类，距离公式为

$$d_i = \sqrt{\sum^n (k_1 - k_2)^2} \quad (5)$$

其中， k_1 和 k_2 分别属于点集 Q 。由于其只考虑点的 x, y, z 信息，因此在分割过程中容易造成过度分割或分割不足问题，影响特征匹配。图 5 中给出了分割不足的情况，为此本文在点云三维坐标基础上加入颜色信息和法线信息进行判断。具体算法如下：

(1) 创建一个空点集 $hull$ ，利用点云凸包算法计算平面 T_{max} 的边界，将计算得到值加入 $hull$ 中；

(2) 创建一个空的索引 I ，计算剩余点在平面 T_{max} 上的投影，如果投影点在边界内，将原始点的索引加入到 I 中，否则跳过；

(3) 为 I 对应的点创建一个快速搜索树，设置一个空的列表 C 和一个空的队列 M ；

(4) 对每一个点 Q_i ，做以点 Q_i 为中心半径 $r < d_{thresh}$ 的球形空间内搜索邻近点集 Q_i^s ，对每一个邻近 $q_j \in Q_i^s$ ，检查 q_i 和 q_j 的距离、色差和法线夹角是否满足给定的阈值，如果小于给定的阈值，则将 q_j 加入队列 M ，当 Q_i 中所有点被处理后，将 M 加入到列表 C 中；

(5) 当所有的点被处理后，程序结束，聚类结果存储于 C 中。

传统分割法和改进分割法得到的结果如图 4~图 6 所示。



图 4 滤波后的剩余点云

由图 5 看出传统的欧式聚类算法只分割了杯子本体，而把杯柄当作另一个聚类，而本文提出的算法较好的将整个杯子分割出来。

2.3 基于 CVFH 全局特征的匹配识别与定位

CVFH 是对视点特征直方图 (VFH) 的一个扩展，在

保证特征旋转缩放不变的情况下提高计算速度，被广泛用于三维点云处理中。本文利用物体聚类分割结果，提取每个聚类的 CVFH 特征，然后与特征数据库中的数据进行匹配，识别出相应的物体，最后计算目标聚类中心。

在特征匹配时，需要首先建立特征数据库。对于每个物体，在特征库中有多个模板特征，模板特征是通过相机在不同视角下拍摄得到，并通过算法从场景中分割出来，并使用 CVFH 算法提取特征。通常为了识别出物体，可以将在线阶段提取的 CVFH 特征描述与特征数据库中的每一个特征描述进行比较，但随着库中特征数量的增多，匹配处理的时间也越长^[15]，同时位置信息无法实时反馈。为了加快匹配速度，本文采用三步法。

(1) 离线训练阶段，将所有特征数据库生成一个搜索树。具体为遍历特征库中所有特征数据，以 CVFH 特征向量的 Chi 平方距离作为相似度量标准，利用 FLANN 快速近似最邻近算法库中的 K-D 树搜索算法来创建快速搜索树序列列表；

(2) 在线识别阶段，以训练阶段的搜索树表为依托，从特征数据库中提取 K 个最邻近聚类特征，在特征空间内，依次计算当前聚类的 CVFH 特征描述子与这 K 个特征的 Chi 平方距离，并对这些距离进行排列，同时存储对应字符标识。通过判断是否满足给定条件来识别出是否是待识别物体，如果顺序表中的最小距离小于阈值则判断识别成功，否则提取下一个聚类的特征进行判断；

(3) 当识别出目标后，遍历目标聚类的整个点云，计算每个点到视觉坐标系原点的距离，取最近点和最远点，计算两点中间值，将该值的信息作为目标在视觉坐标系下的具体位置。

3 实验

3.1 实验平台

本文的实验平台为自主研发的“云”服务机器人 (图 7)，其主要结构为：一个双向轮地盘，两个六自由度的手臂和一个两自由度的视觉系统 (深度摄像机 Kinect)，计算部件是一个配有英特尔酷睿 I7，内存 4G，系统为 Ubuntu12.04 的小型 PC，软件开发工具是开源点云库 (PCL1.7.1)。

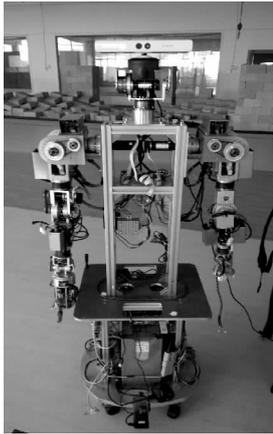


图 7 “云”服务机器人

3.2 实验分析

对于室内机器人, 一般我们考虑两种实验环境, 一种是平面上物体较少的情况即为简单环境; 另一种是平面上放置较多物体, 目标整体可见但是干扰较多的情况。机器人通过自主导航到达桌子附近时, 由于其本体位姿的原因, 很难使目标每次都出现在视场内, 因此需要转动头部来搜寻目标, 并实时把信息反馈给底盘, 以调整自身位姿便于手臂抓取。因此分别在上述两个环境下分析头部在转动和静态两种情况下的识别准确率和单帧处理时间。

首先在简单环境中先考虑转动情况, 如图 8 在桌面上只放一本书, 然后运行识别程序, 并运行控制头部程序使 kinect 左右转动, 实验过程中当物体被绿色框选中时说明目标已被识别, 识别期间记录下 kinect 采集的帧数、有效识别帧数以及处理一帧需要的时间。然后考虑静态情况, 将 kinect 直接正对目标然后运行程序, 同样记录以上数据。

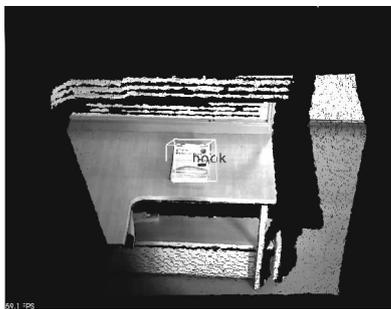


图 8 简单环境

在环境杂乱的桌面上放置一个罐头, 如图 9 所示, 分别进行头部动态和静态的实验, 并记录下数据。

从表 1 中可以看出在简单环境下, 由于水平面上只有一本书, 所以当剔除出平面聚类后只剩下物体聚类, 没有其它干扰项, 因此不管是在动态还是在静态下都达到了 98% 以上的识别率, 而且静态下单帧处理时间只有 0.15 s, 具有较高的实时性。从表 2 中可以看出当识别物体处于复



图 9 复杂环境

杂的桌面环境时识别率有所降低且单帧的处理时间增多, 尤其是头部转动时, 由于场景刷新的原因导致系统需要处理的数据增多, 同时由于水平面上聚类增多, 导致匹配的数目增多, 影响了处理时间, 这也是以后需要改善的地方。

表 1 简单聚类环境

状态	采集帧数	有效识别帧数	识别率/%	单帧处理时间/ms
动态	1536	1511	98.37	287.8
	1813	1786	98.51	286.4
	2245	2213	98.57	287.2
静态	1576	1570	99.61	156.1
	2164	2162	99.90	156.2
	2884	2881	99.89	156.7

表 2 复杂聚类环境

状态	采集帧数	有效识别帧数	识别率/%	单帧处理时间/ms
动态	1212	1002	82.67	477.5
	1756	1397	79.55	487.3
	2118	1788	84.41	482.1
静态	2215	2001	90.33	351.7
	1647	1438	87.31	350.2
	3067	2751	89.69	344.3

表 3 是静态环境下, 文献 [6] 所设计的系统和本文所设计的系统在一些主要过程上的处理时间对比。文献 [6] 在平面处理和物体聚类阶段采用了需要多次迭代的传统方法, 单一的度量准则, 本系统在这两方面进行了改进, 降低了时间复杂度, 提高了识别率。

表 3 各主要过程处理时间对比/ms

系统设计	平面分割	物体聚类	特征计算	匹配识别	位置计算	合计时间
文献 [6]	92.6	8.5	33.9	62.5	3.4	200.9
本文系统	55.3	10.2	33.7	49.1	3.4	151.7

4 结束语

本文设计了一个适用于室内机器人的实时识别与定位系统, 通过对场景点云进行快速有效的分割, 检测出物体

聚类, 并采用基于 CVFH 的特征匹配方法识别目标, 最后计算目标空间位置信息。经过多种情况测试, 该系统在简单环境下有 98% 以上的识别率, 在相对复杂环境下有 85% 以上的识别率, 多次抓取实验结果显示, 算法具有较好的稳定性, 每秒 5 帧的处理速度较好的满足了机器人抓取的实时性要求。由于 CVFH 是一个全局描述子, 依赖于目标分割, 当目标被部分遮挡时难以被整体分割出来, 导致无法识别, 下一步将深入研究局部特征匹配的方法, 使得目标被部分遮挡时也能被识别到。

参考文献:

- [1] Van De Sande K E, Gever T, Snoek C G. Evaluating color descriptors for object and scene recognition [J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2010, 32 (9): 1582-1596.
- [2] Parul Mishra, Nishchol Mishra. Region duplication forgery detection technique based on SURF and HAC [J]. The Scientific World Journal, 2013, 16 (3): 21-29.
- [3] Steder B, Rusu B, Konolige W, et al. NARF: 3D range image feature for object recognition [C] //Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, 2010: 2177-2183.
- [4] Franaszek M, Cheok G S, Witzgall C. Fast automatic registration of range image from 3D imaging system using sphere targets [J]. Automation in Construction, 2009, 18 (3): 265-274.
- [5] Aldoma A, Blodow N, Gossow D, et al. CAD-model recognition and 6DOF pose estimation using 3D Cues [C] //Proceedings of the IEEE International Conference on Computer Vision Workshops, 2011.
- [6] Zhang Qiang, Kong Lingcheng, Zhao Jianghai. Real-time general object recognition for indoor robot based on PCL [C] // Proceedings of the IEEE International Conference on Robotic and Biomimetics, 2013: 651-655.
- [7] Rusu R B, Blodow N, Beetz M. Fast point feature histograms (FPFH) for 3-D registration [C] //Proceedings of the IEEE International Conference on Robotic and Automation, 2009: 3212-3217.
- [8] Rusu R B, Bradski G, Thibaux R, et al. Fast 3D recognition and pose using the viewpoint feature histogram [C] //Proceedings of the 23rd IEEE/RSJ International Conference on Intelligent Robots and Systems, 2010: 2155-2162.
- [9] Rusu R B, Cousins S. 3D is here: Point cloud library (PCL) [C] //Proceedings of the IEEE International Conference on Robotics and Automation, 2011: 1-4.
- [10] Papazov C, Burschka D. An efficient RANSAC for 3-D object recognition in noisy and occluded scenes [C] //Proceedings of the 10th Asian Conference on Computer Vision, 2010: 135-148.
- [11] LI Bao, CHENG Zhiqian, DANG Gang, et al. Survey on normal estimation for 3D point cloud [J]. Computer Engineering and Application, 2010, 46 (23): 1-7 (in Chinese). [李宝, 程志全, 党岗, 等. 三维点云法向量估计综述 [J]. 计算机工程与应用, 2010, 46 (23): 1-7.]
- [12] PAN Guorong, QIN Shiwei, CAI Runbin, et al. Fitted plane automatic extraction algorithm of 3D laser scanning [J]. Journal of Tongji University, 2009, 37 (9): 1250-1255 (in Chinese). [潘国荣, 秦世伟, 蔡润彬, 等. 三维激光扫描拟合平面自动提取算法 [J], 同济大学学报, 2009, 37 (9): 1250-1255.]
- [13] Redono-Cabrera C, Lopez-Sastre R J, Acevedo-Rodriguez J, et al. Surfing the point clouds: Selective 3D spatial pyramids for category-level object recognition [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012: 3458-3465.
- [14] Tombari F, Di Stefano L. Hough voting for 3D object recognition under occlusion and clutter [C] //IPSJ Trans Comput Vis Appl, 2012: 20-29.
- [15] KE Xiang, CHEN Xiaoping, JIN Guoqiang, et al. Real-time object recognition system for indoor service robot [J]. Computer System Application of China, 2013, 22 (10): 84-89 (in Chinese). [柯翔, 陈小平, 靳国强, 等. 一种适用于室内服务机器人的实时物体识别系统 [J], 计算机系统应用, 2013, 22 (10): 84-89.]