一种适用于室内服务机器人的实时物体识别系统①

柯 翔, 陈小平, 靳国强, 王 锋, 郭 群

(中国科学技术大学 计算机科学与技术学院, 合肥 230027)

摘 要: 针对室内服务机器人在实际应用中的需求, 提出一种结合三维点云分割和局部特征匹配的实时物体识别系统. 该系统首先基于三维点云实现快速有效的物体检测, 然后利用物体检测的结果定位物体在彩色图像中的区域, 并采用基于 SURF 特征匹配的方法识别出物体的标识. 实验结果表明, 该系统可较好地满足室内服务机器人物体检测与识别的实时性和可靠性要求.

关键词: 室内服务机器人; 三维点云分割; 局部特征匹配; 实时物体识别

Real-Time Object Recognition System for Indoor Service Robot

KE Xiang, CHEN Xiao-Ping, JIN Guo-Qiang, WANG Feng, GUO Qun

(School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China)

Abstract: For the requirements in practical application of indoor service robot, a real-time object recognition system is proposed which integrates 3D point cloud segmentation and local feature matching. First, it does fast and effective object detection based on 3D point cloud. Then, the areas of objects in color image are located using the result of object detection, and objects are identified using the approach based on SURF feature matching. The experiments show that the system could well meet the real-time and reliability requirements of indoor service robot's object detection and recognition .

Key words: indoor service robot; 3D point cloud segmentation; local feature matching; real-time object recognition

随着科技的进步和人口老龄化问题的加剧, 服务机器人得到了国内外的普遍关注和深入研究. 物体的操作与搬运是服务机器人最重要的基础功能之一, 而一个功能完善的物体操作过程的实现离不开准确、实时的物体检测与识别. 因此, 快速、准确的物体检测与识别是服务机器人研究领域非常具有应用前景的课题. 国际服务机器人标准测试 RoboCup@Home 也专门将物体检测和识别作为了一项挑战测试.

为了能够准确地执行手臂规划和抓取操作, 机器人的物体识别系统不仅需要识别出目标物体的标识, 还需要提供目标物体及其附近障碍物的准确空间位置信息(物体定位). 由于真实环境复杂多变, 在机器人上实现实时可靠的物体识别系统仍是具有挑战性的任务. 一般地, 根据处理数据的类型不同, 机器人的物体识别分为基于二维图像以及基于深度信息的方法. 基于二维图像的物体识别主要利用图像的颜色、纹理、局部特征等, 如文献[1]提出使用颜色直方图法来描述并识别物体, 但该方法受光照影响较大; 文献[2]提出的 SIFT 特征是图像的一种局部特征, 具有尺度缩放、旋转和光照变化的不变性, 在物体识别领域具有广泛应用[3,4]; 文献[5]提出的 SURF 特征是一种快速稳健的局部特征, 与 SIFT 特征相比具有更快的计算速度, 因此更加适用于实时性要求高的场合. 基于二维图像的方法一般较难实现准确的物体定位, 而基于深度信息的方法在识别物体的同时可以给出目标物体准确的空间位置信息, 因此在机器人领域得到了广泛应用. 文献[6,7]提出了使用点特征直方图来表示并识别三维物体的方法, 文献[8]提出了基于深度图像 NARF 特征的物体识别方法. 此类方法均使用深度信息实现物体的检测, 并运用三维几何特征进行物体识别, 可于复杂场景中实现准确的

① 收稿时间:2013-04-01;收到修改稿时间:2013-04-27

物体定位, 但只对几何特征差异较大的物体具有较好的识别效果.

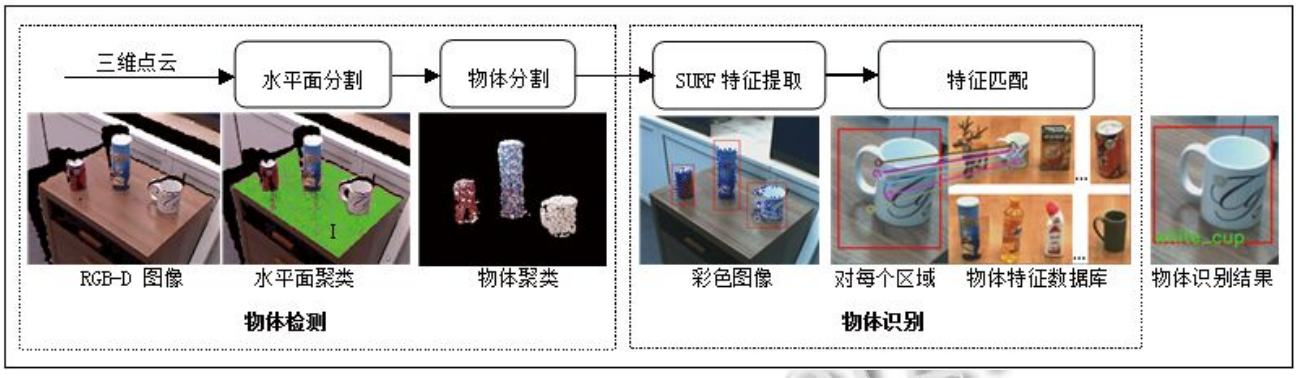


图 1 物体识别系统总体流程

本文提出一种结合三维点云和彩色图像处理的实时物体识别系统, 通过对三维点云的分割实现物体的检测和定位, 并采用基于 SURF 特征匹配的方法实现物体的识别. 实验结果表明该系统在满足实时性的同时具有较高的物体识别率. 在 2012 年 RoboCup@Home 物体识别的相关测试中, 配置了该系统的“可佳”服务机器人表现优异 [1].

1 系统结构

本文的物体识别系统总体流程如图 1 所示, 输入为三维点云和彩色图像, 共分为两步. 第一步实现物体的检测, 由于室内环境中的机器人可操作物体(如饮料、杯子等)通常放置于桌子、架子等水平面上, 我们首先将这些水平面从三维点云中分割出来(如图 1, 水平面分割过程中 RGB-D 图像 [2] 上Ⅰ处绿色标记的区域), 以此为基础再对每个水平面之上的剩余点云进行分割得到相互独立的物体聚类, 这些聚类即检测出的物体. 第二步进行物体识别,首先定位每个物体聚类在彩色图像中的区域, 并提取每个区域的 SURF 特征, 然后与物体特征数据库进行特征匹配, 输出识别的物体标识. 下面详述这两步的实现细节.

2 基于三维点云分割的物体检测

由 RGB-D 摄像头或 TOF(Time of Flight)深度摄像头获取的三维点云, 其数据经过组织后, 在结构上类似于二维图像或二维矩阵, 即每个行列位置对应一个三维点的 x、y、z 坐标. 在此基础上, 本文实现了一种简单、快速有效的物体检测方法, 分为水平面分割和物体分割两个过程.

2.1 水平面分割

从三维点云中分割出水平面, 通常可以采用基于 RANSAC[9]的方法(文献[10]), 即每次随机选取三个点, 计算由这三个点构成的平面模型, 并按照某个设定的距离阈值拟合点云中的其它点. 由此方法分割出的水平面可能包含较多法向量不垂直于水平面的噪声点, 且在处理复杂场景时所花费时间较多, 甚至可能失败. 为了更加准确快速地进行水平面分割, 本文首先计算三维点云中每个点的法向量,然后采用区域增长法直接分割出三维点云中的所有水平面聚类.

对于三维点云 P 中行列位置 (i, j) 处的三维点 p (坐标为 (x_{ij}, y_{ij}, z_{ij}), 记作 $P(i, j)$), 计算该点处两个局部表面切向量的叉乘, 可以得到其法向量 $\vec{n}_p$:

$$\vec{n}_p = \vec{v}_1 \times \vec{v}_2 \quad (1)$$

其中 $\vec{v}_1$ 和 $\vec{v}_2$ 为点 p 的两个局部表面切向量. 它们由以点 p 为中心, 一定大小的正方形像素邻域内近似计算而得. 由于深度传感器获取的数据可能包含噪声, 为了较平滑地计算点 p 的法向量, 邻域的大小由点 p 与坐标原点间的距离来确定, 若距离较小, 则选择较小

[1] 在 clean up 测试中, “可佳”机器人成功识别并顺利抓取到了 4 个物体, 仅次于冠军队伍(成功识别了 5 个物体). 而在 restaurant 测试中, “可佳”是唯一成功在真实酒吧场景中识别并抓取到物体的机器人.

[2] 图 1 中的 RGB-D 图像由 Kinect 获取, 由于其彩色图像分辨率不高, 本文只使用其三维点云数据进行物体检测, 而在物体识别阶段使用具有更高分辨率的彩色相机获取的彩色图像.

的邻域，否则选择较大的邻域. 下面详细介绍法向量 $\vec{n}_p$ 的计算方法.

首先分别求以点 p 的上下左右四个相邻位置点 p_u、p_d、p_l、p_r 为中心，四个边长为 $2k$ 的正方形像素邻域内所有点的三维坐标的平均值 $\bar{p}_u$、$\bar{p}_d$、$\bar{p}_l$、$\bar{p}_r$. 以为例，如图 2 所示，其中阴影部分区域表示以为中心，边长为 $2k$ 的正方形像素邻域，求得该区域内所有点的三维坐标的平均值 $\bar{p}_r$. 为了加快计算的速度，我们事先建立三维点云 P 的积分图像 I，I 中每个行列位置 (m,n) 处的元素值 $I(m,n)$ 为 P 中从左上角位置 $(0,0)$ 到 (m,n) 处的矩形区域内所有点的三维坐标向量的和(即所有点的 x、y、z 坐标分别相加)，按照如下公式迭代求解:

$$I(m,n)=P(m,n)+I(m-1,n)+I(m,n-1)-I(m-1,n-1) \quad (2)$$

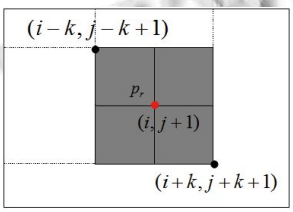


图 2　点的正方形像素邻域

当 $m<0$ 或 $n<0$ 时，$I(m,n)=(0,0,0)$. 因此，不论 k 取何值，只访问积分图像 I 的四个元素即可求出 $\bar{p}_r$:

$$\bar{p}_r=[I(i+k,j+k+1)-I(i-k,j+k+1) -I(i+k,j-k+1)+I(i-k,j-k+1)]/4k^2 \quad (3)$$

采用同样的方法，可分别计算出 $\bar{p}_u$、$\bar{p}_d$、$\bar{p}_l$.

再根据公式(4)、(5)估计点 p 的两个局部表面切向量:

$$\vec{v}_1=(\bar{p}_u-\bar{p}_d)/2 \quad (4)$$

$$\vec{v}_2=(\bar{p}_r-\bar{p}_l)/2 \quad (5)$$

求得切向量之后，由公式(1)可计算出点 p 的法向量.

采用以上方法可得到三维点云 P 中每个点的法向量. 如图 3 所示，上图为一个带有法向量的三维点云，下图为桌面部分的局部放大示意图. 其中白色直线表示各个点的法向量(为了观看清晰，图中只显示了部分点的法向量)，可以看到桌面上的点的法向量方向近似竖直.

由于属于水平面的点的法向量方向近似竖直，因此三维点云中连续且具有近似竖直方向法向量的点可以构成一个水平面聚类. 利用三维点云 P 中所有点的法向量，采用区域增长法可以分割出所有满足一定大小的水平面聚类，算法步骤如下:

1) 遍历三维点云 P，判断每个行列位置处的点的法向量方向是否近似竖直，若是则设置该点的标记为 1，否则设置其标记为 0.

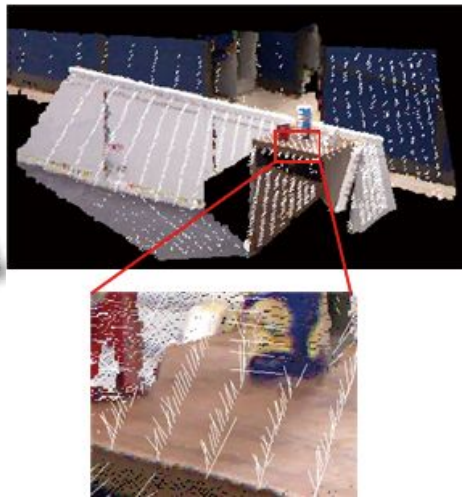


图 3　带有法向量的三维点云及局部放大示意图

2) 顺序扫描三维点云 P，若所有点的标记都为 0，则程序结束. 否则，初始化一空的点集合 T，从第一个标记为 1 的点 p 开始进行区域增长.

3) 将点 p 加入集合 T，并设置点 p 的标记为 0. 分别考虑点 p 的上下左右四个相邻位置的每个点 p_{next}，若 p_{next} 的标记为 1 且 p 与 p_{next} 之间的空间距离小于某一距离阈值，则将 p_{next} 加入队列 q.

4) 若队列 q 非空，则取出队头元素，将其当作点 p，继续步骤 3)；否则转 5).

5) 本次区域增长结束，判断集合 T 大小，若大于某一阈值，则将 T 内的所有点构成的水平面聚类加入水平面集合 S. 转至 2)进行下一次区域增长.

按照以上算法，最终得到水平面分割的结果集合 S，其中每个元素为一个水平面聚类.

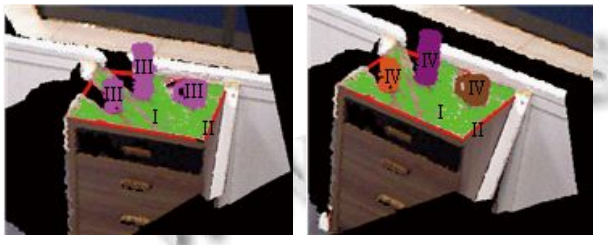
2.2 物体分割

下面，我们分别对水平面分割结果集合 $S=\{T1, T2,\cdots,Tn\}$中的每个水平面 Ti 进行物体分割.

在对水平面 Ti 进行物体分割之前，需要先确定 Ti 的边界以及该水平面 Ti 上的物体的点云. Ti 的边界由 Gramham 算法[11]计算出的包围 Ti 的所有点的二维凸包构成. Ti 上物体的点云由三维点云 P 中所有高于 Ti

的点垂直投影到 *Ti* 后，投影点包含在 *Ti* 边界之内的点构成. 如图 4(a)所示的三维点云中，Ⅰ处绿色部分表示分割出的水平面，Ⅱ处红色多边形表示该水平面的边界，Ⅲ处紫红色部分表示该水平面上的物体的点云.

物体分割的方法与水平面分割类似，首先对三维点云 *P* 中每个行列位置处的点进行标记，将属于水平面 *Ti* 上物体的点标记为 1，其余点标记为 0. 然后采用区域增长法分割出 *Ti* 上相互独立的物体聚类. 图 4(b)给出了(a)中Ⅰ处水平面上的物体的点云分割后的结果，其中Ⅳ处随机彩色部分即表示分割出的相互独立的物体聚类.



(a) 待分割物体的点云 (b) 相互独立的物体聚类

图 4 物体分割示例

3 基于局部特征匹配的物体识别

SURF 特征是图像的一种快速稳健的局部特征，具有一定的尺度不变和旋转不变特性，被广泛应用于图像匹配、物体识别等相关领域[12,13]. 本文利用物体检测的结果定位每个物体聚类在彩色图像中的区域，提取每个区域的 SURF 特征，与物体特征数据库进行特征匹配，识别出每个物体的相应标识.

3.1 特征提取

基于图像 SURF 特征的物体识别，通常有两种特征提取的方式. 一种是提取整幅图像的 SURF 特征，但此种方法计算量大，且图像背景部分的特征对物体识别的效果影响较大；另一种方法是只提取图像中物体区域的 SURF 特征，但事先需要对图像进行分割，当场景复杂时很难实现准确的图像分割. 本文不直接对图像进行分割，而利用物体检测的结果快速准确地定位出物体在彩色图像中的区域，再提取该区域的 SURF 特征.

由于现实环境中机器人可操作的物体的大小在一定范围内，因此不考虑物体检测结果中过小和过大的物体聚类. 对每个大小合适的物体聚类 $Ci(i=1,2,\cdots,n)$，根据透视投影原理计算 *Ci* 的每个点在彩色图像上的投影像素点. 并建立一个包围所有投影点的最小矩形作为物体聚类 *Ci* 在彩色图像中对应的区域 *Ri*，*Ri* 中也包含了图像中物体边界的特征，然后使用 SURF 算法计算出该区域的特征点和相应的 SURF 特征向量.

3.2 特征匹配

在通过特征匹配进行物体识别之前，需要建立物体特征数据库. 对于每个物体，数据库中存储了多个特征模板，其中每一个特征模板为从某个视角拍摄的只包含该物体的一幅图像中，使用 SURF 算法计算出的特征点和相应的 SURF 特征向量.

为了识别出彩色图像中每个区域 $Ri(i=1,2,\cdots,n)$的物体标识，可将 Ri 与数据库中的每个物体的每个特征模板分别进行特征匹配，但这种方法耗费时间较长. 为了提高物体识别速度，本文采用一种两步特征匹配的方法.

第一步，将数据库所有特征模板的特征点，组成一个集合，称为数据库特征点集合. 对 *Ri* 的每个特征点，以 SURF 特征向量间的欧氏距离作为特征点的相似性度量标准，使用 FLANN 快速近似最近邻算法库[14]中的随机 K-D 树搜索算法，在数据库特征点集合中查找最近邻特征点. 依据最近邻特征点所属的特征模板，统计出 *Ri* 与每个特征模板匹配到最近邻特征点的个数，选择匹配个数最多的前 *m* 个作为候选特征模板集，并按由多到少的顺序排列.

第二步，将 *Ri* 与候选特征模板集中的特征模板依次进行特征匹配. 若 *Ri* 与某个特征模板匹配成功，则该特征模板所属的物体的标识即为 *Ri* 的识别结果，并不再继续与其它特征模板进行匹配；若 *Ri* 与 *m* 个特征模板都匹配失败，则 *Ri* 的识别结果为未知物体. *Ri* 与特征模板的匹配算法如下:

1) 以 SURF 特征向量间的欧氏距离为相似性度量标准，使用随机 K-D 树搜索算法查找 *Ri* 的每个特征点 p_i 在特征模板中的最近邻特征点 p_{n1} 和次近邻特征点 p_{n2}，若最近邻距离和次近邻距离的比值小于某一阈值，则将 p_i 与 p_{n1} 组成的匹配点对加入候选匹配点对集合 *St*.

2) 对候选匹配点对集合 *St*，使用 RANSAC 算法迭代地计算出 *Ri* 与特征模板之间的最佳仿射变换模型，并删除 *St* 中不满足变换模型的匹配点对. 若最终 *St* 中的匹配点对个数大于某一设定的阈值，则判定 *Ri* 与该特征模板匹配成功，否则认为匹配失败.

4 实验

4.1 实验平台

本文的实验平台为“可佳”服务机器人(如图 5 所示). 其主要硬件结构为: 一个双向轮底盘、一个六自由度的手臂和一个具有两自由度云台的视觉硬件系统(深度相机 Kinect, 高分辨率彩色相机 POINT GREY GRAS-14S5C-C)等. 其计算部件是一台配置有 Intel Core i7-2760QM 处理器、4GB 内存, 并运行 Linux 系统的笔记本.

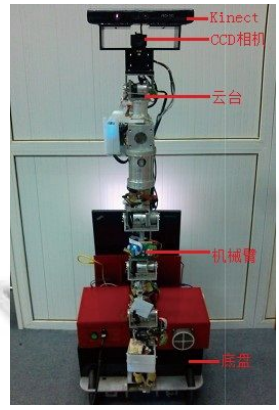


图 5 “可佳”服务机器人

4.2 实验数据

本文的实验数据分为物体特征模板数据以及实验测试数据, 这些数据均使用“可佳”服务机器人采集. 其中物体特征模板数据用于构建物体特征数据库, 测试数据用于测试系统的物体识别效果. 实验共设置了 30 个物体, 分成 A、B、C 三组, 每组 10 个,图 6 给出了部分物体.

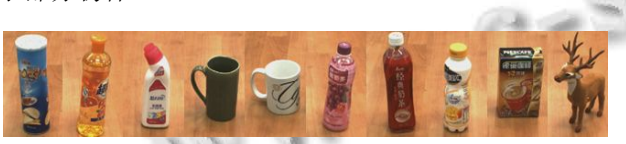


图 6 实验部分物体示例

物体特征模板数据: 将每个物体放在背景干净的桌面上, 围绕物体一周每隔约 45° 视角, 由机器人的高分辨率彩色相机拍摄一幅只包含该物体的图像, 共拍摄 8 幅. 对每幅图像, 提取其特征点和相应的 SURF 特征向量作为该物体的一个特征模板.

实验测试数据: 为了测试本文的物体识别系统对场景的适应性, 我们在餐桌、水槽和书架三个不同室内场景中, 对 A、B、C 三组物体分别独立采集一组测试数据, 称为 A'、B'、C'组. 采集每组物体的测试数据时, 将该组每个物体放到每个场景中至少 8 次, 且每个场景中每次同时放置 2-6 个物体, 由机器人上的 Kinect 获取的三维点云和高分辨率彩色相机拍摄的彩色图像构成一个测试实例. 图 7 给出了三个不同场景的测试示例. 表 1 给出了在 A'、B'、C'三组测试数据中, 物体在每个场景中出现的次数以及在三个场景中出现的总次数.

(a) 餐桌场景　(b) 水槽场景　(c) 书架场景

图 7 三个不同场景的测试示例

表 1 物体在三组测试数据中出现的次数

测试数据	餐桌场景	水槽场景	书架场景	总次数
A' 组	93	86	88	267
B' 组	99	91	93	283
C' 组	98	90	87	275

4.3 实验与分析

为了测试不同规模数据库对系统物体识别效果的影响, 本文在不同实验数据集上进行了三次实验. 第一次实验数据库中存储 A 组物体的特征模板数据, 对 A'组测试数据进行测试; 第二次实验数据库中存储 A 和 B 两组物体的特征模板数据, 对 A'和 B'两组测试数据进行测试; 第三次实验数据库中存储 A、B 和 C 三组物体的特征模板数据, 对 A'、B'和 C'三组测试数据进行测试. 表 2 给出了三次实验的物体识别结果, 表 3 给出了三次实验中系统在单个测试实例上各个阶段平均运行时间以及总的平均运行时间.

表 2 三次实验的物体识别结果

物体特征数据集/测试数据集	餐桌场景	水槽场景	书架场景	总识别率
A/ A'	97.84%	96.51%	97.73%	97.38%
A+B/ A'+B'	94.79%	94.35%	93.37%	94.18%
A+B+C/ A'+B'+C'	88.62%	89.51%	89.92%	89.33%

从表 2 可以看出, 在第一次和第二次实验中, 当数据库只有 10 个和 20 个物体的特征模板数据时, 系统在三个不同场景中的物体识别率和总的物体识别率都达到了 90%以上. 在第三次实验中, 当数据库有 30

个物体的特征模板数据时, 系统在三个场景中总的物体识别率达到了 89.33%, 仍然具有较高的物体识别率. 不难发现, 随着数据库规模的增大, 系统的物体识别率有所下降, 主要是由于物体数目增多时, 数据库中的不同物体的相似特征点数目也相应增加, 物体识别时相互之间的干扰造成的.

表 3 单个测试实例上的平均运行时间(单位: 毫秒)

物体特征数据集/测试数据集	水平面分割	物体分割	特征提取	特征匹配	总时间
A/ A'	56.3	10.9	49.1	35.8	152.1
A+B/ A'+B'	55.7	11.3	48.7	39.5	155.2
A+B+C/ A'+B'+C'	55.2	10.8	49.6	44.9	160.5

由表 3 可知, 当数据库有 30 个物体的特征模板数据时, 系统在单个测试实例上的平均运行时间为 160.5 毫秒, 平均达到 6 帧每秒的速度, 具有良好的实时性. 同时可以看出, 该系统在不同实验数据集上的水平面分割、物体分割、特征提取的平均时间都基本保持不变, 然而随着数据库规模的增大, 特征匹配的时间也相应增加, 原因主要是物体识别时, 需要与更多的物体的 SURF 特征进行特征匹配.

5 结语

实时的物体检测和识别是机器人视觉领域一个具有挑战性的任务. 本文设计并实现了一种适用于室内服务机器人的物体识别系统, 通过三维点云分割实现快速有效的物体检测和定位, 采用 SURF 特征匹配的方法实现准确的物体识别. 在真实室内环境多个场景中的实验结果以及 RoboCup@Home 物体识别相关测试的结果共同表明了该系统有效可行, 可较好地满足室内服务机器人物体检测和识别的实时性和可靠性要求. 在本文的基础上, 下一步我们将尝试采用多种特征融合的物体识别方法, 如结合图像纹理特征和三维几何特征, 进一步扩展机器人可识别的物体范围.

参考文献

1 Swain MJ, Ballard DH. Color indexing. International Journal of Computer Vision, 1991, 7(1): 11–32.

2 Lowe DG. Object recognition from local scale-invariant features. International Conference on Computer Vision, 1999, 7: 1150–1157.

3 Romea AC, Torres MM, Srinivasa S.The MOPED framework: Object recognition and pose estimation for manipulation. International Journal of Robotics Research, 2011, 30(10): 1284–1306.

4 Gordon I, Lowe DG. What and where:3D Object Recognition with Accurate Pose. In: Ponce J, Hebert M, Schmid C, Zisserman A, eds. Toward category-level object recognition. Lecture Notes in Computer Science. Springer, 2006, 4107: 67–82.

5 Bay H, Ess A, Tuytelaars T, van Gool L. SURF: Speeded up robust features. Computer Vision and Image Understanding (CVIU), 2008, 110(3): 346–359.

6 Rusu RB, Blodow N, Beetz M. Fast point feature histograms (FPFH) for 3D registration. Proc. of the IEEE International Conference on Robotics and Automation(ICRA). Kobe, Japan. 2009.

7 Rusu RB, Bradski G, Thibaux R, Hsu J. Fast 3D recognition and pose using the viewpoint feature histogram. Proc. of the 23rd IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS). Taipei, 2010.

8 Steder B, Rusu RB, Konolige K, Burgard W.NARF:3D range image features for object recognition. Workshop on Defining and Solving Realistic Perception Problems in Personal Robotics at the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems(IROS). Taipei. 2010.

9 Fischler MA, Bolles RC. Random sample consensus:A paradigm for model fitting with applications to image analysis and automated cartography. Comm. of the ACM, 1981, 24: 381–395.

10 Nuechter A, Hertzberg J.Towards semantic maps for mobile robots.Journal of Robotics and Autonomous Systems (JRAS), Special Issue on Semantic Knowledge in Robotics, 2008:915–926.

11 Graham RL.An efficient algorithm for determining the convex hull of a finite planar set. Information Processing Letters, 1972, 1(4): 132–133.

12 张锐娟,张建奇,杨翠.基于 SURF 的图像配准方法研究.红外与激光工程,2009,38(1):160–165.

13 范新南,丁朋华,刘俊定,张学武.融合灰度和 SURF 特征的红外目标跟踪.中国图象图形学报,2012,17(11):1376–1383.

14 Muja M, Lowe DG. Fast approximate nearest neighbors with automatic algorithm configuration. International Conference on Computer Vision Theory and Applications (VISAPP). 2009.