移动机器人基于三维激光测距的室内场景认知

庄 严 [1] 卢希彬 [1] 李云辉 [1] 王 伟 [1]

摘 要 研究了移动机器人在室内三维环境中的场景认知问题. 室内场景框架具有结构化特性, 而室内多样化的物体则难以进行模型化表述. 本文利用区域扩张算法进行平面特征的提取, 并根据平面属性及其相互间的空间关系, 完成室内场景框架的辨识. 为了借鉴图像处理领域的物体识别方法, 本文提出一种基于 Bearing Angle 模型的激光测距数据表述方法, 从而将三维点云数据转换为二维 Bearing Angle 图. 同一类物体中的个体形态具有多样性, 同时观测视角也导致激光测距数据的显著差异. 针对这些问题, 采用一种基于 Gentleboost 算法的有监督学习方法, 并利用物体碎片及其相对于物体中心的位置作为特征, 从而完成室内场景中的物体认知. 利用室内场景框架辨识结果在 Bearing Angle 图中进行天棚、地面、墙壁、房门等区域的标记, 并利用所产生的语义信息去除错误的认知结果, 从而有助于提高识别率. 利用实际机器人平台所获得的实验结果验证了所提方法的有效性.

关键词 三维激光测距, Bearing Angle 图, Gentleboost, 室内场景认知

DOI 10.3724/SP.J.1004.2011.01232

Mobile Robot Indoor Scene Cognition Using 3D Laser Scanning

ZHUANG Yan[1] LU Xi-Bin[1] LI Yun-Hui[1] WANG Wei[1]

Abstract This paper mainly studies the indoor scene cognition problem for mobile robots. The structure of an indoor scene is assumed to be structured, while indoor objects are in a variety of forms, which are difficult to be represented by specific descriptive models. In our work, planes are extracted from 3D laser data using regional expansion algorithm, and the properties as well as the relationship of these planes are used for the indoor structure identification. In order to adopt digital image processing algorithm to implement object detection, a bearing angle model is used to represent laser point clouds, so that 3D laser scanning data can be converted to 2D bearing angle image. It is difficult to detect a large number of different classes of objects in cluttered indoor scenes, especially when the mobile robot acquires the 3D laser scanning data in different locations and angles of view. An approach based on gentleboost algorithm is proposed for multi-class object detection, which takes the fragments and the location with respect to the object center as the generic features for object detection. As the result of indoor structure identification, the specific region for ceiling, floor, wall and door can be labeled in the bearing angle image. With the help of known semantic information, the false object detection results can be eliminated effectively. Experiment results implemented on a real mobile robot show the validity of the proposed method.

Key words 3D laser scanning, bearing angle image, gentleboost, indoor scene cognition

移动机器人对其工作环境的有效感知、辨识与认知, 是其进行自主行为优化并可靠完成所承担任务的前提和基础. 移动机器人利用传感器数据所提取的环境特征, 能帮助机器人完成定位与导航等行为, 但并不能保证其可有效实现自主环境理解等高级智能行为. 场景认知问题就是在对环境空间认知的基础上, 形成对场景中物体的理解以及对客观场景的解释, 从而指导机器人完成更高层次的智能行为, 如推理、决策以及基于认知结果的自主环境交互.

收稿日期 2010-11-02 录用日期 2011-03-02
Manuscript received November 2, 2010; accepted March 2, 2011
国家自然科学基金 (61075094, 61035005) 和中央高校基本科研业务费专项资金 (DUT11ZD201) 资助
Supported by National Natural Science Foundation of China (61075094, 61035005) and Fundamental Research Funds for the Central Universities of China (DUT11ZD201)
1. 大连理工大学信息与控制研究中心 大连 116024
1. Research Center of Information and Control, Dalian University of Technology, Dalian 116024

如何实现场景中物体的有效分类与识别是移动机器人场景认知的核心问题. 基于视觉图像处理技术来进行场景的认知是该领域的重要方法[1−3]. 但移动机器人在线获取的视觉图像质量受光线变化影响较大, 特别是在光线较暗的场景更难以应用. 与视觉传感器相比, 三维激光测距具有精度高且不易受光线影响等特点, 因而近些年引起很多国内外学者的关注, 并开展了一系列的相关研究. Himmelsbach 等利用栅格的思想分割物体, 然后在原始激光数据上提取特征, 并使用支持向量机 (Support vector machine, SVM) 分类器对物体进行分类[4]. Anguelov 等完整地描述了一种基于 Markov 随机场的物体分类方法[5]. 该方法将整幅场景视为一个 Markov 随机场, 每一个激光点视为一个结点, 在激光点上提取特征. 相比较, 文献 [6] 中虽然也采用 Markov 随机场模型来对物体进行认知, 但是该方法首先利用激光点的邻域关系从激光数据

中提取多边形集, 然后, 以此作为结点进行物体的分类处理. 上述方法虽然在环境模型中引入场景语义信息, 但并未深入涉及推理与学习等认知领域核心问题, 只能应用于简单场景的认知, 对复杂场景中具有多元化的多类物体认知有一定局限性.

考虑在实际场景认知应用中, 机器人所处位置与视角会发生显著变化, 这也造成在不同次激光扫描中由激光点云所表示的物体形态会有显著变化, 而且由于角度差异也会造成不同物体之间的相互遮挡. 针对这种复杂场景中多形态多类别物体的辨识与认知问题, 基于模型匹配的方法无法很好地完成场景和物体的识别任务, 所出现的认知混淆现象无法保证可靠的认知率. 在模式识别与图像处理领域, 有监督的学习方法是解决本文研究问题的重要手段[7−8]. 借鉴上述研究工作, 并为了对机器人所获取的三维激光测距数据进行有效的简化, 本文提出了一种将三维点云数据转化为二维 Bearing Angle (BA) 图, 并利用 Gentleboost 算法来提高室内场景物体认知准确率的方法. 虽然 Nüchter 等在研究中也将获取的三维激光数据转化为二维深度图, 并使用支持向量机从这些二维图片中训练并识别出物体[9], 但与 BA 图相比, 深度图无法充分反映环境的局部细节信息, 因而对处理形态及尺寸变化显著的物体具有明显的局限. 此外, 与基于纯视觉图像进行有监督学习的物体认知方法进行比较, 本文所提方法具有如下几方面优势: 1) 由于本文研究的是室内场景认知问题, 而室内场景一般可视为封闭的三维空间, 因此, 可从原始三维激光点云中利用区域扩张算法提取室内场景框架信息 (包括墙壁、天棚、地面等框架元素), 并可在生成的 BA 图中对属于不同框架元素的区域进行标记. 本文基于 Gentleboost 算法, 并通过训练 BA 图样本建立起的碎片特征字典来完成室内物体 (如椅子、办公桌、沙发、显示器等元素) 的认知. 由于训练样本数目的局限及多形态多物体认知问题本身固有的困难, 仍会出现一定数量的错误认知结果; 但这些错误结果经常会出现在测试 BA 图样本已标记的场景框架元素区域, 因此, 可以利用语义关系直接去除错误认知结果 (如椅子不能出现在天棚、墙壁等场景框架区域), 从而有效提升认知率. 2) 基于视觉图像进行有监督学习的物体认知方法, 通常只能利用矩形框在图像中圈出物体所在区域, 而无法进一步精确给出该物体的轮廓. 由于 BA 图中像素与三维激光点云的激光点具有单一对应关系, 因此, 可以将 BA 图中获得的认知结果矩形框区域还原成三维激光测距数据, 并进一步利用聚类算法分割出属于该物体的精确激光点云. 3) 根据精确分割出的物体激光点云数据, 可以利用统计方法计算出该物体的几何属性 (如其几何中心、空间尺寸等参数), 并利用这些属性进一步确定物体的认知结果. 通过对室内场景框架信息以及室内物体有效认知, 有助于构建室内环境的语义地图, 这将有效提高机器人与环境的交互能力.

1　基于三维激光测距的室内场景框架信息提取

1.1　三维激光测距系统及其模型

本文激光数据采集系统由激光测距仪、控制激光测距仪旋转的电机、搭载的云台、上位机控制软件以及电源等构成 (见图 1 (a)). LMS200 激光测距仪垂直安放在靠蜗轮蜗杆传动的精密电动旋转云台上, 通过控制器编程控制旋转云台转动, 可以获取旋转范围为 0° ∼ 360° 的三维全景激光数据. 其中, LMS200 激光测距仪扫描范围 180°, 角度分辨率 0.5°, 通讯模块传输速率为 500 KBd.

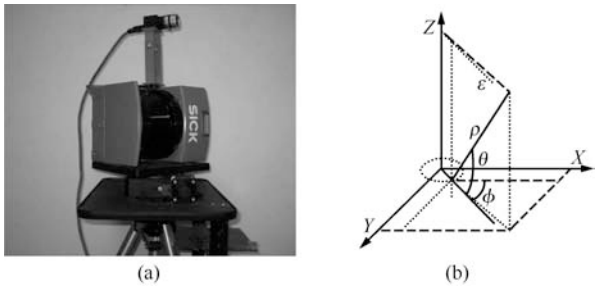


图 1　3D 激光测距系统 (a) 和坐标关系示意图 (b)

Fig. 1　3D laser range finder (a) and the diagram of coordinate relations (b)

由激光测距仪获取的数据为极坐标形式, 为了计算方便, 需要转换到直角坐标系下 (见图 1 (b)), 转换公式如下:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \cos\theta\cos(\varphi-\varepsilon) & \cos\varphi \\ \cos\theta\sin(\varphi-\varepsilon) & \sin\varphi \\ \sin\theta & 0 \end{bmatrix} \begin{bmatrix} \rho \\ c \end{bmatrix} \tag{1}$$

其中, ρ 为激光器光心到被测物体表面的距离; θ 为激光束在扫描平面上的偏向角度; φ 为电控旋转台旋转的角度; c 为电机旋转中心到激光束发射中心投影到水平面上的距离, $c \approx 10\,\text{mm}$; ε 为 LMS 激光扫描平面与两中心点所在直线的夹角, $\varepsilon \approx 4°$.

将激光数据按照扫描顺序保存到二维容器中:

$$P(i,j) = (x,y,z), \quad i = 0,\cdots,M-1, \quad j = 0,\cdots,N-1 \tag{2}$$

其中, $p(i,j)$ 表示第 i 行、第 j 列个激光点, (x,y,z) 表示激光点的空间坐标值, M 表示电机旋转方向激光行数, N 表示每一行激光扫描的激光点的个数.

1.2 基于区域扩张的平面拟合

在室内场景中, 墙壁、地面、天棚等场景框架包含大量平面特征. 因此, 可以从三维激光数据中提取平面, 并利用平面属性及平面之间的空间关系来完成室内场景框架认知. 作为室内场景认知的基础, 平面特征的提取直接影响到后面框架认知的效果. 国内外文献中已经提出了多种从三维激光点云中提取平面特征的算法[10−11]. 本文借鉴 Stamos 等[11] 所使用的方法进行局部邻域平面的拟合, 然后, 将局部平面合并得到完整的平面. 由于激光数据存储在二维容器中, 每个激光点有 8 个邻居 (边界除外), 直接反映了激光数据在空间的邻域关系. 定义一个 $k \times k$ 的邻域, 将邻域内的激光点进行平面拟合. 可以根据实际情况取不同的值, 以图 2 为例, 两个矩形区域均为 4×4 的邻域.

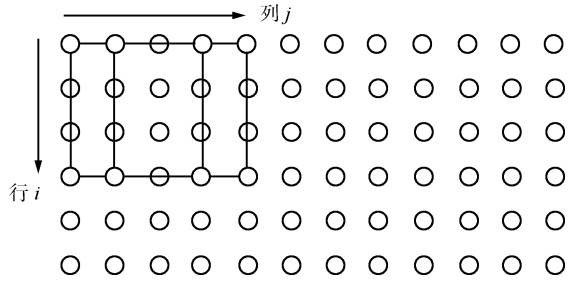


图 2 邻域平面拟合示意图

Fig. 2 The diagram of neighboring plane fitting

每个邻域包含 $p_1, p_2, \cdots, p_n$ 共 n 个激光点, 设其重心为 w, 有

$$w = \frac{1}{n}\sum_{i=1}^{n} p_i \tag{3}$$

设拟合后平面的单位法向量为 $\boldsymbol{n}$, 拟合的目标是使得下式取得最小值,

$$e = \sum_{i=1}^{n}((p_i - w)^{\mathrm{T}} * \boldsymbol{n})^2 \tag{4}$$

实际应用中为 e 设置一个阈值 E, 当 $e < E$ 时认为该邻域是一个平面.

由于拟合后的 $k \times k$ 邻域平面只是完整平面的片段, 相邻的邻域平面需要进一步合并以提取出完整平面, 合并的条件是两邻域平面的法向夹角足够小. 为了方便后续室内场景框架的认知, 同时还计算出平面的重心、法向量和平面方程等属性, 并保存其对应点云数据的索引号.

1.3 室内场景框架信息提取

在平面特征提取完毕之后, 一个完整的平面可能由于物体遮挡或者激光数据缺失等因素被断开 (例如已提取的多个平面可能归属于同一面墙), 因此, 需要对这些平面进行合并. 平面合并条件包括如下两个方面 (如图 3 所示): 条件 1 为待合并两平面 A 和 B 的法向量 L_a 和 L_b 的夹角 α 小于一定阈值; 条件 2 为平面 A 的重心 O 到平面 B 的距离 d 小于一定阈值. 室内场景中的墙壁、天棚、地面、门等结构化场景元素, 均可以用大面积的平面特征来表述. 以提取的平面特征为基础, 借助平面之间的空间关系与逻辑关系, 以及平面自身的属性信息来判断其属于哪种框架元素. 参照室内框架认知示意图 4, 可按照如下方法完成室内场景框架的认知.

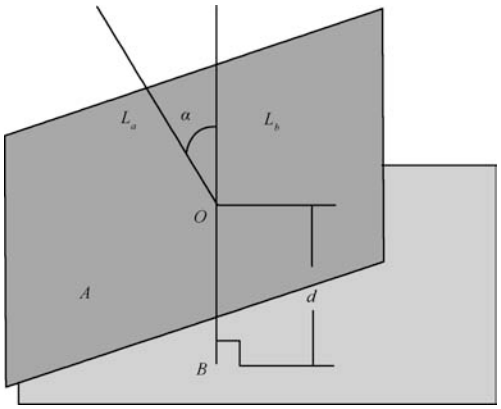


图 3 平面特征合并示意图

Fig. 3 The diagram of planar features merging

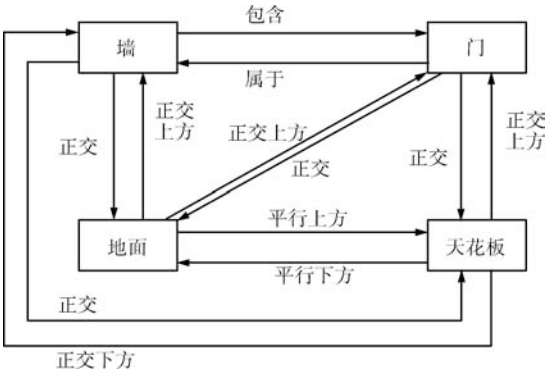


图 4 室内框架认知示意图

Fig. 4 The diagram of indoor framework recognition

步骤 1. 找出法向与 Z 轴平行的平面, 设平面的重心高度为 z, 按照 z 的大小顺序将这些平面排序.

步骤 2. 将 z 值最小且包含点云数量超过一定阈值的平面标记为地面.

步骤 3. 同理, 将 z 值最大且包含点云数量超过一定阈值的平面标记为天棚. 如果查找失败, 则分别从底部和顶部向 z 的中值趋近查找. 另外, 按照一般

$$BA_{i,j}=\arccos\frac{\sqrt{x_{i,j}^2+y_{i,j}^2+z_{i,j}^2}-\sqrt{x_{i+1,j+1}^2+y_{i+1,j+1}^2+z_{i+1,j+1}^2}\cdot\cos d\phi}{\sqrt{(x_{i,j}-x_{i+1,j+1})^2+(y_{i,j}-y_{i+1,j+1})^2+(z_{i,j}-z_{i+1,j+1})^2}} \tag{5}$$

$$d\phi=\arcsin\frac{\sqrt{a^2+b^2+c^2}}{\sqrt{x_{i,j}^2+y_{i,j}^2+z_{i,j}^2}\sqrt{x_{i+1,j+1}^2+y_{i+1,j+1}^2+z_{i+1,j+1}^2}} \tag{6}$$

建筑的特点, 要求天棚距离地面的高度大于 2.5 m.

步骤 4. 找出法向与 Z 轴垂直的平面, 求取其最低点和最高点. 将满足平面顶部接近天棚, 底部接近地面的平面标记为墙壁.

步骤 5. 根据经验知识, 将满足如下条件的墙壁元素标记为门: 1) 该平面的高度 $H=1.8\,\mathrm{m}\sim2.2\,\mathrm{m}$; 2) 该平面的底部接近于地面.

2　基于 BA 图和 Gentleboost 算法的室内物体认知

2.1　BA 图生成及场景框架信息标记

为了简化三维激光点云数据的像素邻域关系, 可将三维激光点云数据转成二维图像, 从而可以借鉴图像处理方法来实现物体的认知. 本文采用了一种改进的 BA 图模型进行三维激光点云的二维图形化处理. 在本文所采用的改进模型中, BA 图的像素值被定义为如图 5 所示的特定两个相邻激光测距点之间的夹角值 $BA_{i,j}$. 在文献 [12] 中 BA 图被首次提出并用来完成摄像机的外标定, 但其计算的是第 i 组扫描中相邻两点 $(P_{i,j},P_{i,j+1})$ 所构成的夹角; 为了兼顾激光测距信息在横纵两个方向的角度变换, 在本文则计算第 i 组扫描中第 j 个激光点和第 $i+1$ 组扫描中第 $j+1$ 个激光点 $(P_{i,j},P_{i+1,j+1})$ 所构成的夹角 (如图 5 所示).

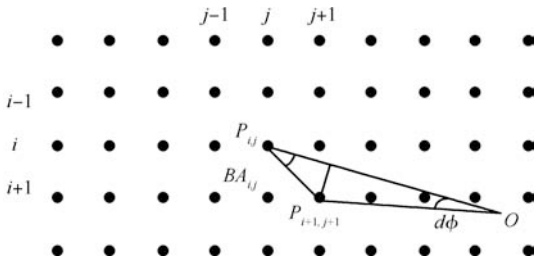


图 5　两个相邻激光测距点所构成的夹角

Fig. 5　The angle between two adjacent laser points

在本文应用中, 构建一幅 BA 图像的式子如式 (5) 和式 (6), 其中, $d\phi$ 是两条激光束之间的夹角, $a=x_{i,j}y_{i+1,j+1}-y_{i,j}x_{i+1,j+1}$, $b=y_{i,j}z_{i+1,j+1}-z_{i,j}y_{i+1,j+1}$, $c=z_{i,j}x_{i+1,j+1}-x_{i,j}z_{i+1,j+1}$. 由于图像中的像素与激光点云具有一一对应关系, 所以可以基于第 1 节已提取的室内场景天棚、地面、墙壁等框架信息, 对 BA 图所表述场景的对应框架信息进行标记. 图 6 所示为任意选取的一个室内场景的 BA 图, 图中天棚、地面、墙壁等框架信息分别被标记出, 而图 6 中间未被标记的部分是需要进行深入认知的室内物体所在区域.

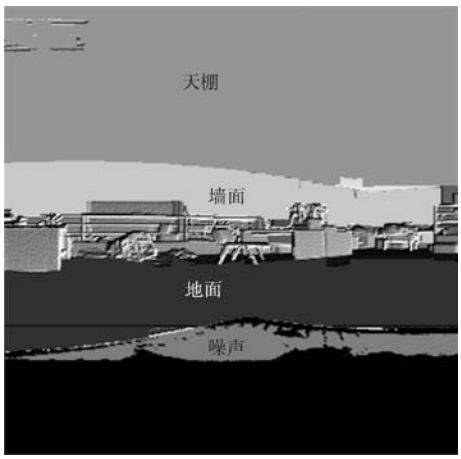


图 6　根据场景框架信息进行标记的 BA 图

Fig. 6　BA image labeled according to the scene framework information

2.2　Gentleboost 算法原理

考虑室内物体的多样性及其所处场景的复杂多变, 本文采用 AdaBoost 的一种变体 Gentleboost 算法[13] 进行室内物体的认知. AdaBoost 符合下面的叠加模型:

$$F(x)=\sum_{m=1}^{M}f_m(x) \tag{7}$$

其中, x 为特征值, M 是训练的轮数, $f_m(x)$ 为弱分类器, $F(x)$ 为强分类器. Gentleboost 算法在每轮的训练中优化下面的成本函数:

$$J=\sum_{i=1}^{N}w_i(z_i-f_m(x_i))^2 \tag{8}$$

其中, N 是训练样本 (特征值) 的数目, w_i 为样本权重, z_i 为样本标签 (±1). 使这个成本函数最小化取

决于弱分类器 f_m. 本文定义的弱分类器为一个简单的函数 $f_m(x) = a\delta(x^j > \theta) + b\delta(x^j \le \theta)$, $f_m(x)$ 由 4 个参数 $[a, b, \theta, j]$ 来确定, x^j 是待训练的第 j 个特征, θ 是阈值, δ 是指示函数, a 和 b 为回归参数, 可按如下式获得:

$$a = \frac{\sum_{i=1}^{N} w_i z_i \delta(x_i^j > \theta)}{\sum_{i=1}^{N} w_i \delta(x_i^j > \theta)} \tag{9}$$

$$b = \frac{\sum_{i=1}^{N} w_i z_i \delta(x_i^j \le \theta)}{\sum_{i=1}^{N} w_i \delta(x_i^j \le \theta)} \tag{10}$$

x_i^j 是第 j 个特征所对应的第 i 个观测特征值. 每个弱分类器可以被看作是只有一个节点的退化决策树 (回归树), 所以可以像学习一个决策树的节点那样找出最优的回归树: 对每一个特征 x^j 对应的观测特征值进行排序, 并找出所有可能的阈值 θ, 给定了 j 和 θ 以及相应的 a 和 b, 选择使式 (8) 最小的 $[a, b, \theta, j]$, 并且利用这个弱分类器更新强分类器 $F(x) \leftarrow F(x) + f_m(x)$. 最后更新每个训练样本的权重 $w_i \leftarrow w_i \exp(-y_i f_m(x_i))$ 并归一化 $w_i \leftarrow \frac{w_i}{\sum_{i=1}^{N} w_i}$, 这样增加误分类的样本权重 $(y_i f_m(x_i) < 0)$, 减小正确分类的样本权重 $(y_i f_m(x_i) > 0)$. 如此循环 M 次, 可得到一组比较优的弱分类器 $f_1, f_2, \cdots, f_M$, 并进一步确定强分类器 $F(x) = \sum_{m=1}^{M} f_m(x)$.

2.3 基于 Gentleboost 算法的室内物体认知

利用文献 [8, 14] 中使用的碎片作为特征, 建立一个碎片特征字典. 从所有室内场景的 BA 图样本中, 随机选择 n 张作为训练样本. 在实验中发现一些经过滤波器滤波的碎片特征能够更有效地进行物体认知, 所以选择 3 种滤波器来对图像进行滤波, 它们分别是 Sobel 算子沿行方向的 3×3 模板、沿列方向的 3×3 模板和拉普拉斯算子的 3×3 模板, 加上原图片总共有 4 种滤波器:

$$s_1 = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}, \quad s_2 = \begin{bmatrix} -1 & 0 & 1 \\ 2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

$$s_3 = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}, \quad s_4 = \begin{bmatrix} 1 \end{bmatrix} \tag{11}$$

碎片是从图像物体实例上任意提取的包含部分物体信息的图像片段, 这些正方形片段的边长在 $19 \sim 25$ 个像素之间任意变化, 和每一个提取的碎片相关的是一个二值化图像, 表示碎片在图像中相对于物体中心的位置. 用 $x(s, p, g)$ 来表示一个碎片特征, s 表示滤波器, p 表示经过滤波的碎片, g 表示碎片相对于物体中心的位置, 部分碎片特征如图 7 所示.

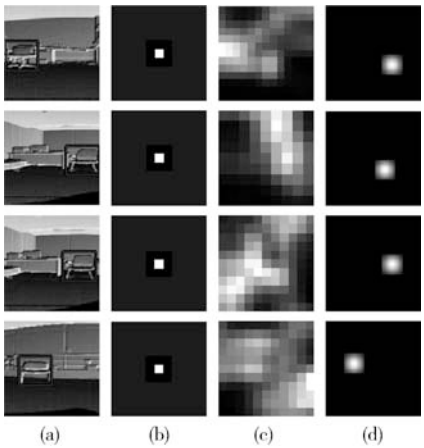


图 7 (a) ∼ (d) 分别为样本图片、滤波器、碎片、碎片相对于物体中心的位置

Fig. 7 (a) ∼ (d) are specimen images, filters, patches, and the locations relative to the object center, respectively

建立碎片字典后, 利用每一个碎片特征 $x_j(s, p, g)$ 计算特征值, 其计算式如下:

$$h_j(I, x, y) = [|I * s_j| \otimes p_j] * g_j \tag{12}$$

这里 I 为输入图像, "$*$" 表示卷积运算, "$\otimes$" 表示相关 (Cross-correlation) 运算, $h_j(I, x, y)$ 为输出图像. 因为一幅图像有成千上万个像素, 所以会产生大量的训练样本 (特征值), 为了减少训练样本的数目, 只选择 $h_j(I, x, y)$ 上所有位置的一个子集. 在物体的几何中心位置提取正训练样本, 在背景位置随机提取 30 个负样本 (如图 8 所示). 根据计算的特征值经过 M 轮训练, 得到 M 个最优弱分类器.

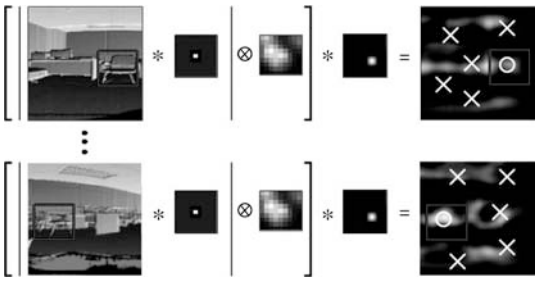


图 8 计算特征值的过程 (输出图像中圆圈表示正样本特征值 (其标签 $z_i = 1$); 交叉线表示负样本特征值 (其标签 $z_i = -1$))

Fig. 8 The caculation process of eigenvalues (the circle in output images shows positive sample eigenvalues (the label $z_i = 1$); the cross line shows negative sample eigenvalues (the label $z_i = -1$).)

在测试 BA 图片上进行物体认知, 对于每个弱分类器分别利用式 (12) 和函数 $f_m(x) = a\delta(x^j > \theta) + b\delta(x^j \le \theta)$ 计算特征值和检测值, 并把所有的检测值求和形成强的分类输出, 又称为分类分数. 对分类分数大于 0 的二值化图像进行 Hamming 低通滤波, 然后求出所有其分数大于周围八邻域分数的位置坐标, 可得到局部区域分数最大点的坐标 (x, y). 如果两个局部区域分数最大的点坐标之间的距离小于一定的阈值, 取这两个坐标的几何中心位置为分数最大的坐标. 确定了中心点的坐标和分数值, 画出以中心点为中心的边界框. 根据 BA 图已提取的室内场景框架信息, 可以有效地去除检测错误的物体, 从而提高识别率. 例如, 座椅和显示器都不会出现在天棚和墙面上, 显示器也不会出现在地面上. BA 图中检测物体的矩形框内部包含不属于物体实例的像素, 利用 BA 图中像素与三维激光点云的一一对应关系, 在三维激光数据中利用 K-means 聚类方法[15] 进行聚类, 可精确获得属于物体的三维激光点云数据. 聚类后根据属于物体的三维激光数据的几何中心进一步去除认知错误的物体, 例如座椅的中心应在距地面 0.4 m ∼ 0.8 m 这一范围内.

3　实验结果

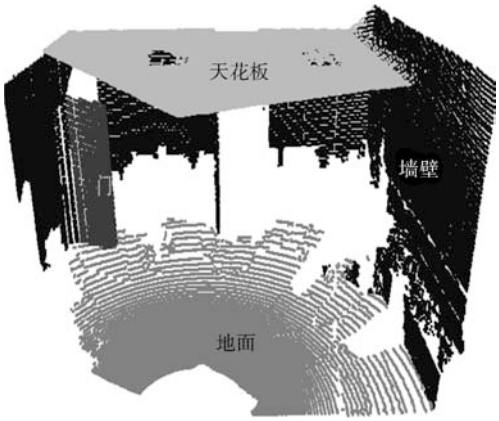
本实验中所使用的移动机器人平台为美国先锋公司的 Pioneer3-DX 移动机器人, 其配备有课题组自主研发的三维激光测距系统 (如图 1 左图所示). 为了更好地分析和显示移动机器人三维场景建模与自主场景认知的实验结果, 课题组自主研发了专用的处理软件 3D SMART.

3.1　室内场景框架信息提取实验结果

室内框架信息的提取实验选取了课题组所在的大连理工大学创新园大厦 6 楼 A0613 房间的一处场景 (如图 9 (a) 所示). 机器人在该处场景中, 激光的扫描范围是 150°. 竖直方向上每组激光包含 361 个点, 共有 596 组, 包含的激光点个数为 596 × 361. 图 9 (b) 显示了该场景的框架信息提取效果 (注意这里没有显示物体部分). 可以看出, 室内的各个框架被清晰地提取出来.

(a) 室内场景图片

(a) Vision image of an indoor scene



(b) 室内场景框架认知实验结果

(b) Result of an indoor scene framework recognition

图 9　室内场景框架认知实验

Fig. 9　Indoor scene framework cognition experiment

3.2　室内物体认知实验结果

以教师办公室、研究生办公室、机器人实验室等多个不同房间为实验环境, 随机采集了 204 组三维激光测距数据, 并均转换为 BA 图. 上述场景中所包含的主要物体为椅子、办公桌、沙发、电脑显示器等物体. 利用第 2 节所提出的基于有监督学习方法, 将上述 204 幅 BA 图分为训练样本和测试样本两类. 训练样本共计 92 幅 BA 图, 其中, 46 幅用于提取碎片特征并建立碎片特征字典, 余下的 46 幅用来计算特征值, 具体数据见表 1.

表 1　不同物体用于建立碎片特征字典和用于计算特征值的训练样本数目

Table 1　The sample numbers of different objects used to establish patch dictionary and calculate the eigenvalues

BA 图片数目 (幅)	椅子	办公桌	沙发	显示器	合计
用于建立碎片特征字典	12	10	12	12	46
用于计算特征值	12	10	12	12	46

余下的 112 幅 BA 图作为测试样本. 图 10 为随机给出的 4 组不同物体 (图 10 中 1～4 行分别为椅子、办公桌、沙发、显示器) 的认知结果, 在 BA 图中均用矩形框圈出. 由于测试样本中的椅子、办公桌、显示器这三种物体的形态均有显著差异, 移动机器人进行三维激光数据采集时所处视角的显著差异, 以及被其他物体的遮挡, 致使物体显示在 BA 图中的形态也不尽相同, 这些都给机器人的自主认知带来了挑战. 图 10 所给出的认知结果表明, 以 BA 图所表述的一定量场景为训练和测试样本, 利用 Gentleboost 算法和建立碎片特征字典, 能够对办公室环境中的桌椅、沙发、显示器等常见物体进行有效的认知.

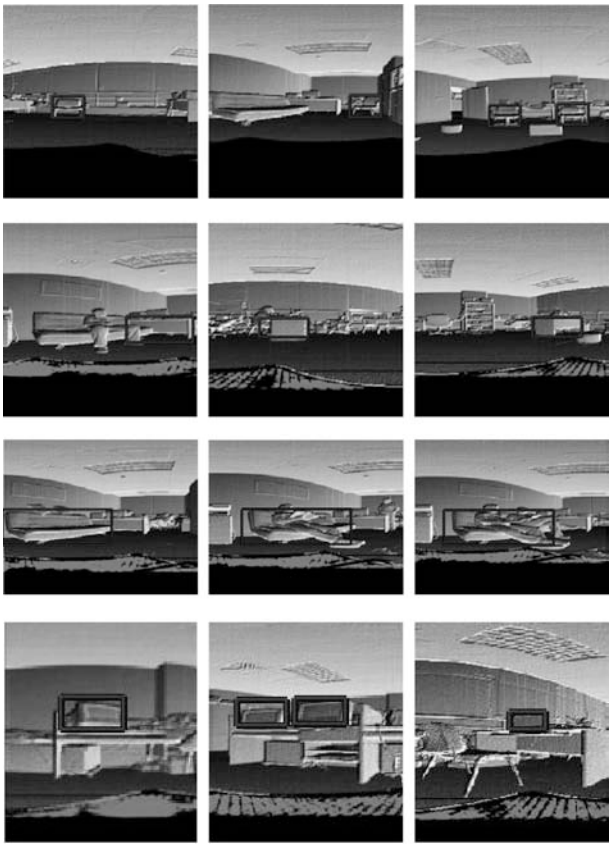


图 10 随机给出的 4 种不同物体的认知结果, 第 1～4 行分别为椅子、办公桌、沙发、显示器

Fig. 10 Recognition results of four different objects shown randomly (Lines 1～4 are chairs, desks, sofas, monitors, respectively.

在测试样本的 BA 图中, 标识被认知物体的矩形框内部包含有不属于该物体的像素. 在传统的基于视觉图像的物体认知中, 如果要精确去除不属于该物体的图像像素是非常困难的. 但由于二维 BA 图中像素与原始三维激光点云具有单一映射关系, 所以可以将该矩形框区域的 BA 图像素还原为三维激光点云, 并利用聚类方法获取属于该物体的三维激光点云数据. 图 11 所示为以三维激光点云所表示的某一室内场景认知结果.

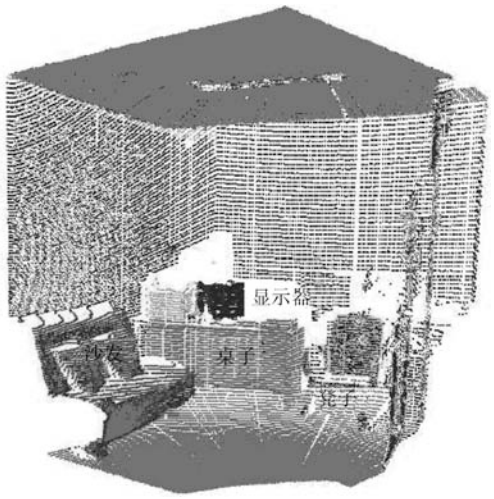


图 11 激光点云表述的室内场景中多种物体的认知结果

Fig. 11 Indoor objects recognition results shown in 3D laser point clouds

3.3 实验结果分析

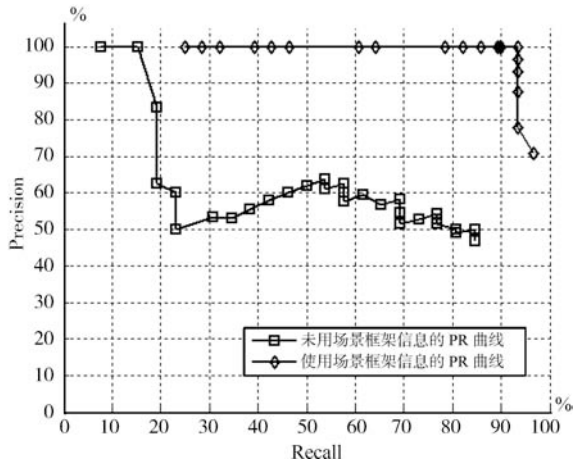
本文所提方法与基于纯视觉图像进行有监督学习的物体认知方法相比较, 一个明显优势是可利用从原始激光点云中提取出的天棚、地面、墙壁、门等场景框架元素的语义信息, 剔除出现在 BA 图中与框架元素对应区域内的错误物体认知结果. 图 12 中给出的三组室内物体的 BA 图认知结果中, 灰色矩形框所示区域为正确认知物体, 而白色矩形框所示区域为错误认知物体. 图 12 中从左到右分别为座椅、显示器和办公桌的认知结果, 而出现在地面、墙壁和天棚的相应认知结果是错误的, 能够有效去除.

图 12 根据 BA 图中场景框架语义信息去除错误的认知结果

Fig. 12 The wrong recognition results removed using the semantic information of the scene framework

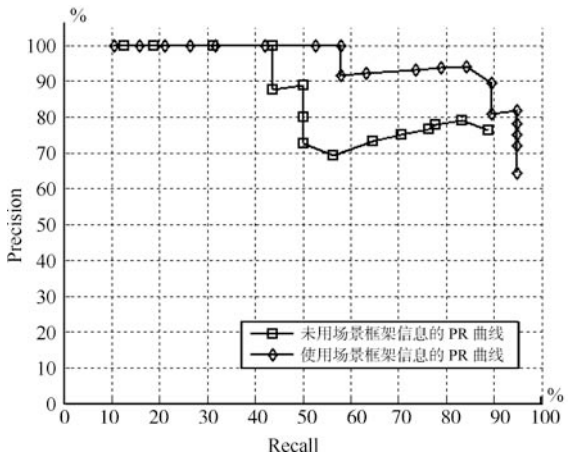
正确率–召回率 (Precision-Recall, PR) 曲线能够很好地反映物体检测和认知的性能. 召回率 (Recall) 也称为查全率, 表示正确认知的物体数占场景中待识别物体总数的比率 (记为 $RObjTrue/ObjSum$); 正确率 (Precision) 也称为

查准率, 表示正确认知的物体数占进行实际认知的物体总个数的比率 (记为 $RObjTrue/RObjSum$); 其中, $RObjTrue$ 为正确认知的物体个数, $ObjSum$ 为场景中待识别的物体的总个数, $RObjSum$ 为进行实际认知的物体总个数. 这两个指标分别度量检索效果的查全和查准两方面, 忽略任何一个方面都有失偏颇. 图 13 分别为椅子和显示器认知结果的 PR 曲线, 其横坐标为召回率, 纵坐标为正确率. 分析图 13 所示 PR 曲线可知, 使用场景框架信息进行物体认知可有效提升认知的正确率, 同时又不影响物体认知的召回率.



(a) 认知椅子的正确率 – 召回率曲线

(a) PR curves for chair recognition



(b) 认知显示器的正确率 – 召回率曲线

(b) PR curves for monitor recognition

图 13　椅子和显示器的正确率 – 召回率曲线

Fig. 13　PR curves for chair and monitor recognition

4　结论

本文主要研究了移动机器人室内场景认知的问题. 针对室内环境的特点, 通过框架信息的认知, 三维激光点云到 BA 图的转换, 场景框架信息在 BA 图中的标记, 基于 Gentleboost 算法的有监督学习, 错误认知物体的去除和基于聚类的物体激光点云细化分割等过程, 最终实现对室内场景的整体认知. 框架信息的认知主要应用了平面特征提取技术, 而针对具有多态性的室内多物体认知问题, 由于充分利用了已获取的场景框架元素语义信息, 可有效提高物体认知率. 如何与其他学习方法相结合, 并降低认知算法的技术复杂度, 将是下一步研究工作重点.

References

1 Sande K E A, Gevers T, Snoek C G M. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(9): 1582−1596

2 Felzenszwalb P F, Girshick R B, McAllester D, Ramanan D. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(9): 1627−1645

3 Vernaza P, Lee D D. Scalable real-time object recognition and segmentation via cascaded, discriminative Markov random fields. In: Proceedings of the IEEE International Conference on Robotics and Automation. Anchorage, USA: IEEE, 2010. 3102−3107

4 Himmelsbach M, Luettel T, Wuensche H J. Real-time object classification in 3D point clouds using point feature histograms. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. St. Louis, USA: IEEE, 2009. 994−1000

5 Anguelov D, Taskarf B, Chatalbashev V, Koller D, Gupta D, Heitz G, Ng A. Discriminative learning of Markov random fields for segmentation of 3D scan data. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE, 2005. 169−176

6 Agrawal A, Nakazawa A, Takemura H. MMM-classification of 3D range data. In: Proceedings of the IEEE International Conference on Robotics and Automation. Kobe, Japan: IEEE, 2009. 2003−2008

7 Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE, 2001. 511−518

8 Torralba A, Murphy K P, Freeman W T. Sharing visual features for multiclass and multiview object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **29**(5): 854−869

9 Nüchter A, Hertzberg J. Towards semantic maps for mobile robots. *Robotics and Autonomous Systems*, 2008, **56**(11): 915−926

10 Surmann H, Lingemann K, Nuchter A, Hertzberg J. Fast acquiring and analysis of three dimensional laser range data. In: Proceedings of the International Fall Workshop Vision, Modeling and Visualization. Stuttgart, Germany: AkaGmbH, 2001. 59−66

11 Stamos I, Allen P K. 3-D model construction using range and image data. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hilton Head Island, USA: IEEE, 2000. 531−536

12 Scaramuzza D, Harati A, Siegwart R. Extrinsic self calibration of a camera and a 3D laser range finder from natural scenes. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. San Diego, USA: IEEE, 2007. 4164−4169

13 Friedman J, Tibshirani R, Hastie T. Additive logistic regression: a statistical view of boosting. *The Annals of Statistics*, 2000, **28**(2): 337−374

14 Ullman S, Sali E, Vidal-Naquet M. A fragment-based approach to object representation and classification. In: Proceedings of the 4th International Workshop on Visual Form. London, UK: Springer, 2001. 85−102

15 Macqueen J. Some methods for classification and analysis of multivariate observations. In: Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. Berkeley, USA: University of California Press, 1967. 281−297

庄 严 大连理工大学自动化系副教授. 主要研究方向为机器人导航、探索、自主环境建模与环境认知. 本文通信作者.
E-mail: zhuang@dlut.edu.cn
(**ZHUANG Yan** Associate professor in the Department of Automation, Dalian University of Technology. His research interest covers mobile robot navigation, exploration, autonomous environment modeling and cognition. Corresponding author of this paper.)

卢希彬 大连理工大学信息与控制研究中心硕士研究生. 主要研究方向为室内场景认知.
E-mail: lxb19861226@yahoo.com.cn
(**LU Xi-Bin** Master student at the Research Center of Information and Control, Dalian University of Technology. His main research interest is indoor scene cognition.)

李云辉 大连理工大学信息与控制研究中心硕士研究生. 主要研究方向为室内场景认知.
E-mail: niceapplebanana@163.com
(**LI Yun-Hui** Master student at the Research Center of Information and Control, Dalian University of Technology. Her main research interest is indoor scene cognition.)

王 伟 大连理工大学信息与控制研究中心教授. 主要研究方向为预测控制、机器人学及智能控制.
E-mail: wangwei@dlut.edu.cn
(**WANG Wei** Professor at the Research Center of Information and Control, Dalian University of Technology. His research interest covers predictive control, robotics, and intelligent control.)