

文章编号: 1001-0920(2010)01-0001-07

## 基于单目视觉的移动机器人导航算法研究进展

夏庭锴<sup>a</sup>, 杨明<sup>b</sup>, 杨汝清<sup>a</sup>

(上海交通大学 a. 机器人研究所, b. 自动化系, 上海 200240)

**摘要:** 基于单目视觉的移动机器人导航的研究, 涵盖了机器视觉、模式识别和多目标跟踪多个领域. 其算法框架不仅成功应用于移动机器人导航, 还为目标检测、识别与跟踪领域的研究提供了可供参考的模型. 该综述将以算法发展历史为脉络, 结合一些典型系统, 通过对关键技术和算法结构的分析比较, 总结算法本身的发展前景和由此发展起来的可供相关研究参考的算法框架.

**关键词:** 移动机器人导航; 机器视觉; 模式识别; 目标跟踪

**中图分类号:** TP242.6 **文献标识码:** A

### Progress in monocular vision based mobile robot navigation

XIA Ting-kai<sup>a</sup>, YANG Ming<sup>b</sup>, YANG Ru-qing<sup>a</sup>

(a. Research Institute of Robotics, b. Automation Department, Shanghai Jiaotong University, Shanghai 200240, China. Correspondent: XIA Ting-kai, E-mail: xiatingkai2004@sjtu.edu.cn)

**Abstract:** The monocular vision based mobile robots navigation system is studied based on three corresponding areas, machine vision, pattern recognition and multiple targets tracking, which provides a useful algorithm framework for object recognition and tracking except for the navigation. The development history of the existing algorithms and some typical systems are reviewed. Then some promising technologies and algorithm structures are analyzed. Finally, the prospective of the algorithm is discussed and the algorithm framework that maybe useful for related research fields is summarized.

**Key words:** Mobile robot navigation; Machine vision; Pattern recognition; Object tracking

### 1 引言

室外移动机器人导航技术的基本任务是, 通过传感器给出移动机器人与环境的相对位姿, 为机器人进行局部路径规划. 在已有的导航技术中, 基于视觉的方法由于其丰富的传感器信息和最接近人类的感知方式, 成为最具有研究前景的方法之一. 在视觉方法中, 基于单目视觉的算法由于在鲁棒性和效率之间达到了较好的平衡, 且只需要相对廉价的设备, 从而获得了最为广泛的研究. 到目前为止, 这类算法的发展大致分为 3 个阶段: 第 1 阶段研究主要针对结构化环境的应用, 其中小范围环境中通常使用基于地图的方法, 但基本停留在实验室阶段; 针对公路环境的应用则不创建全局地图, 大多依赖地面标线特征进行局部相对位姿估计, 主要面向智能交通领域的应用, 如商用的道路偏移警报系统, 考虑到应用背景, 通常还包括障碍物检测<sup>[1]</sup>或系统状态监测的

内容<sup>[2]</sup>. 第 2 阶段的研究开始转向复杂的结构化道路、城市环境和非结构化环境(野外环境). 机器学习理论, 更复杂的滤波技术以及传感器融合技术被引入该检测系统, 传感器层面和算法层面的融合成为这个时期研究的重点. 但同时, 由于面临的问题更为发散和复杂, 未象第 1 阶段的研究那样形成一个“标准”的算法框架. 第 3 阶段研究的基本思想是“基于单幅图像的三维重构”, 试图通过分析单幅图像在多尺度下的特征, 以基于机器学习的方法来重构三维场景, 最终分析出移动机器人周围环境的“可通行区域”. 这个研究思路将原先由距离传感器(如超声测距仪、激光雷达等)完成的任务改为完全用视觉方法来实现, 以包含障碍物信息的局部地图代替结构化环境中的“道路”概念, 为复杂环境下的移动机器人视觉导航应用提供了更有吸引力的方法, 是排爆、消防机器人实现全自主导航的最有前景的方法之一.

收稿日期: 2009-01-07; 修回日期: 2009-04-14.

基金项目: 上海科委浦江人才计划项目(07pj14055); 教育部博士点基金项目(20070248097).

作者简介: 夏庭锴(1980—), 男, 江苏如皋人, 博士生, 从事移动机器人导航系统的研究; 杨汝清(1944—), 男, 江苏江阴人, 教授, 博士生导师, 从事特种机器人、智能车辆等研究.

## 2 算法概论

基于地图的方法在室内移动机器人的导航中获得了较为成功的应用,如基于 SIFT (Scale-Invariant Feature Transform) 特征的视觉 SLAM (Simultaneous Localization And Mapping)<sup>[3]</sup>. 但室外环境的复杂性、不确定性和大范围的特点使得基于地图的方法在可靠性和地图表示等方面受到很大限制. 因此,早期的基于视觉的室外移动机器人导航研究大多采用无地图方法,研究背景主要针对结构化环境,包括高速公路车辆巡航和工厂环境的自动导引车 (AGV),如 VaMP<sup>[4]</sup>, GOLD<sup>[5]</sup>, RALPH<sup>[6]</sup>, LANA<sup>[7]</sup> 等系统. 算法流程通常包括特征提取、环境参数估计和系统状态跟踪 3 个部分,而且一般采用如下假设来简化问题:

- 1) 环境地面具有标线且规格已知或部分已知;
- 2) 环境地面平坦且模型已知(形状、范围等);
- 3) 兴趣区域假设.

到 21 世纪初为止,最典型和最完整的工作是由德国联邦国防军大学 Dickmans 教授领导的研究小组完成的 EMS-Vision (Expectation-based Multifocal Saccadic Vision) 视觉系统<sup>[4,8]</sup>和意大利帕尔马大学的 GOLD 系统<sup>[5]</sup>.

针对结构化环境的应用对地面标线和环境结构有较强的依赖性,但室外移动机器人不可避免地要面临非结构化的环境,此时由于地面纹理复杂甚至没有明显的道路区域,上述面向结构化道路的算法将不再适用. 甚至在结构化环境中,由于特征提取算法对噪声的敏感,阴影、遮挡问题也可能会使算法失效. 因此,在上述针对结构化环境系统被提出的同时,也诞生了一些针对非结构化环境的系统,其中较具代表性的有 ALVINN 系统<sup>[11]</sup>.

21 世纪以来,研究的重点基本集中在局部方法的改进和传感器融合技术的系统化,这既得益于新的数学方法的发展,也受益于硬件水平的不断提高. 在面向结构化环境的应用上,研究的重点集中于更复杂的路况,如交叉路口检测、无地面标线或地面标线状况不理想的情形<sup>[9]</sup>,但整体思路仍遵循传统框架. 在非结构化环境方面,由于 2004 年和 2005 年面向沙漠环境的 DARPA 大挑战赛的开展,这方面的研究获得了较系统化的进展<sup>[10-12]</sup>. 由于面向非结构化环境的系统本身面临的问题相当复杂,实用系统一般采用多传感器方案,基于单目视觉的导航算法作为整个传感器系统的一个组成部分,离单独完成导航任务还有相当的距离. 此外,随着滤波理论的发展,环境模型参数估计与跟踪领域的研究也有较大进步,如基于多模型的道路跟踪算法<sup>[9,13]</sup>,基于粒子

滤波的道路检测算法和数据融合框架等<sup>[2,14,15]</sup>. 但滤波技术本身对环境模型仍然有较强的依赖,并不能从根本上解决移动机器人的“绑架”问题,所以对于完全依靠视觉方法的导航系统而言,能最终解决问题的是对图像结构的“彻底理解”. 即在没有很强的先验知识、其他传感器信息以及历史数据的情况下,从图像获取可靠的环境信息,直至重建整个三维场景. 如果把环境检测看成物体识别问题,可以使用基于统计学习理论的方法进行可通行区域检测,在人脸检测、交通标志识别领域广泛使用的支持向量机 (SVM), AdaBoost (Adaptive Boost) 等机器学习方法已被广泛应用于此类算法中<sup>[16,17]</sup>. 理论上,重建三维场景需要依靠立体视觉,但立体视觉本身面临视场有限、设备复杂以及算法可靠性等问题,实际应用受到一定限制. 近年来,一些研究者在单幅图像的理解方面做出了很多努力,已经可以初步创建出三维场景信息<sup>[18-20]</sup>,开辟出视觉导航领域一个很有前景的研究方向.

## 3 结构化环境中的典型方法

20 世纪 80 年代,受限于硬件水平和不成熟的算法,结构化环境相对简单的背景使得研究者们能更有效地提出基本问题,以及更快地建立实际可用的实验系统,从而成为当时的主要研究对象.

EMS-Vision 系统<sup>[8,21]</sup>是结构化环境应用研究领域中最具有代表性的一套系统,其核心思想为“4D 模型方法”. 即空间 (3D) 与时间 (1D) 在扩展卡尔曼滤波 (EKF) 理论框架下的结合,适用于所有地面有导引线或明显的可通行区域边界的应用场合.

4D 方法由 2 个主要模块合成:1) 特征检测与参数估计模块,通过沿水平方向搜索灰度图像中“暗-明-暗”的灰度变化,检测具有一定方向和宽度的地面标线或自然边界,并利用检测出的边缘特征估计道路参数;2) 系统状态预测与模型更新模块,预测结果(机器人与道路的相对位姿)将新的检测兴趣区域限制在上一帧图像处理结果的一定范围内,以提高图像处理的效率和部分解决参数空间先验分布

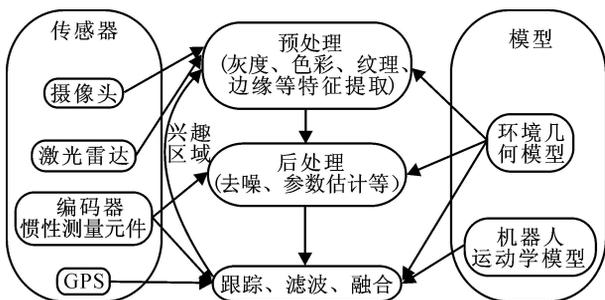


图 1 结构化环境下的典型检测算法框架

非高斯的问题。

这个时期,面向实际系统的算法为了提高可靠性,已开始进行传感器融合的研究(主要是距离传感器和惯性导航元件),其基本框架可用图 1 来概括。这类算法包含两个主要步骤:特征检测和环境几何模型估计。

### 3.1 特征检测

边缘特征是在早期的研究中,使用最广泛的环境特征。结构化环境中,清晰且规格已知的地面标线以及已知的环境范围信息,使基于边缘特征的可通行区域检测在保证效率的同时也具有较高的鲁棒性。而且,边缘特征直接反映了可通行区域的形状,非常便于使用回归方法得到环境的几何参数,适用于工厂、公路等具有规则形状的环境。Sobel, Canny 等常用的边缘检测算子可用于进行地面标线的边缘提取,是早期视觉导航系统的常用方法<sup>[5, 8, 22, 23]</sup>。通用的边缘检测算法对噪声比较敏感,所以通常需要对图像进行预处理和后处理以得到更理想的边缘检测结果。例如,建筑物、杂物以及其他移动物体造成的阴影会产生很多“假边界”,是结构化环境中通行区域检测算法可靠性的主要威胁。研究者们提出了很多解决方案,最基本的方法是利用环境的先验知识,如已知的标线规格或道路宽度;其次利用基于灭点检测的方法,对地面标线或自然边界进行一致性检测;最后利用启发式算法找到平行的边界<sup>[24]</sup>。此外,对图像进行快速傅立叶变换,通过图像的频域特征来识别地面标线也可在一定程度上提高鲁棒性<sup>[25]</sup>。最广泛使用的仍是基于滤波技术的方法,即根据前一帧的检测结果确定当前帧的检测区间,通过不断调整感兴趣区域的方法提高检测速度和鲁棒性<sup>[26]</sup>。

还有一些检测算法游离于主流算法之外,这类算法的基本思想是把地面标线的检测看成轨迹跟踪或物体识别问题。比如一种基于生物进化的方法,受到蚁群寻找食物行为的启发,将经过二值化后的图像上可能代表边界的像素看作“食物”,通过“食物”对“蚂蚁”(当前搜索点)的吸引力(像素距离),改变搜索路径的权值。反复此过程,对图像进行多次搜索,从而选择一条最佳的可能路径作为最终的检测结果<sup>[27]</sup>。在另一些系统中,地面标线被视为运动物体的轨迹,从图像下方开始,所有地面标线上的像素都成为运动物体轨迹的“历史数据”,因而地面标线检测问题就变成了运动物体轨迹跟踪问题。可通行区域边界在图像上的投影通常可采用二次曲线近似描述,从而可使用常规的滤波算法对这些“轨迹”进行跟踪,如基于卡尔曼滤波的方法<sup>[28]</sup>和基于粒子滤

波的方法<sup>[29]</sup>。使用这类方法需要解决滤波器初始化的问题,即找到“轨迹”的初始位置。文献[28]中使用小范围的 Hough 变换找到图像底部的地面标线,文献[29]中则从图像底部开始向上逐行扫描,每行图像数据作为观测值,若干行以后,粒子滤波将收敛到正确结果。

### 3.2 环境几何模型估计与跟踪

特征提取的结果是一系列代表了地面标线或可通行区域边界的特征点,为了给出规划模块能够理解的信息,还需对这些特征点进行回归分析,得到环境模型的几何参数(通常采用直线模型或二次曲线模型)。Hough 变换或其改良版本是这类算法中常用的参数估计方法。在可通行区域检测这个实时性要求较高的应用中,通常采用的是随机 Hough 变换(RHT)<sup>[30-32]</sup>和多分辨率 Hough 变换(Multi-resolution Hough Transform)<sup>[33]</sup>,以避免传统 Hough 变换在待估计参数较多时运算量大和需要量化参数空间所造成的速度劣势。为保证检测结果的可靠性,通常采用 Kalman 滤波器对系统状态进行跟踪。系统状态一般包括两部分信息:移动机器人位姿参数(与规划路径的横向偏差和航向角偏差)和环境模型的参数(曲率、宽度等)。文献[26]中使用了一个比较特殊的参数估计方法,它构造的系统状态是一个 20 维的向量

$$X_d = (u_{1L}, u_{2L}, \dots, u_{10L}, u_{1R}, u_{2R}, \dots, u_{10R})^T,$$

其中  $u_{iL}$  和  $u_{iR}$  ( $i = 1, 2, \dots, 10$ ) 分别表示从图像中抽取的 10 行数据中第  $i$  行的左右地面标线上点的横坐标(图像坐标系下)。这些点将图像分成 18 个子区域,区域的初始左右边界由训练生成的置信区间确定。地面标线检测分别在這些子区域中进行。地面标线可能不连贯,所以并非所有的子区域都能检测到标线,因此在检测到地面标线的子区域中可通过基于 Kalman 滤波的方法进行迭代,不断收缩置信区间,最终让  $X_d$  收敛于实际的标线。该方法极大地提高了有地面标线的结构化环境中可通行区域检测的可靠性,但其严格的先验假设使它的应用范围受到限制。

文献[32]中使用的自适应随机 Hough 变换(ARHT)是结构化环境参数估计比较有代表性的方法。该方法考虑到 Hough 变换中参数空间的维数不能太大,所以为了提高运算效率,对环境模型作了降维处理。原始模型为

$$u = k/(v - hz) + b \times (v - hz) + vp. \quad (1)$$

其中: $u$  和  $v$  为图像坐标, $k$  与地面坐标系下的道路曲率成比例, $vp$  为道路切线方向, $b$  为摄像机相对于道路边线的位移, $hz$  为地平线在图像坐标系下的纵

坐标. 对式(1)进行降维处理(两边对  $v$  求导), 有

$$u = \frac{2k}{(v - hz)} + \left(\frac{du}{dv}\right)(v - hz) + vp. \quad (2)$$

其中:  $du/dv = -f_g(u, v), f_g(u, v)$  是由 Sobel 算子得到的边缘梯度方向.

式(2)通过微分变换回避了在随机 Hough 变换中估计道路宽度  $b$ . 该做法的依据是: 由地面标线和其他道路边界产生的所有边缘特征拟合出的边界曲线, 在地面坐标系下是互相平行的(具有不同的参数  $b$  和相同的参数  $k, vp$ ), 从而进行随机 Hough 变换时不需要再对特征点(边缘)进行分类(即判断特征点是否属于同一曲线). 这一点对提高随机 Hough 变换的效率至关重要. 在从“特征点集”中抽取点进行参数计算时, 作者借鉴粒子滤波的思想对每个特征点进行了权重评估, 具有较大边缘梯度幅值的点具有较大的权重, 而权重大的特征点将有更大的概率被抽中, 用于计算待估计参数, 从而提高整个算法的效率.

在参数估计中同样有一些较为特殊的方法, 如受复杂背景下动态物体跟踪算法的启发, 有的研究者将结构化道路上的地面标线作为一个整体进行检测和跟踪, 使用样条曲线作为地面标线的模型, 把地面标线检测和参数估计问题转化为运动物体跟踪问题<sup>[28, 34, 35]</sup>.

#### 4 半结构化环境中的导航方法

面向结构化环境的道路检测, 完成了移动机器人可通行区域检测基本方法的研究, 由此面向城市环境的应用被提到研究日程上. 城市环境道路相对于结构化环境的应用, 其道路状况更加复杂, 需要考虑多种道路模型以及交叉路口等情形, 算法层面和传感器层面的数据融合技术成为这个时期实用系统的主要研究方向.

在基于单目视觉的导航方法的算法层面, 可利用的特征除结构化道路检测中广泛采用的边缘点外, 还有纹理、色彩等. 早期的检测算法忽略这些有效信息的原因, 很大程度上是由于在很长的时间里一直没有一个高效的算法框架来融合这些信息, 以满足实时应用的需要. 随着粒子滤波理论在工程领域越来越广泛的应用, 这个问题得到了一定程度上的解决.

粒子滤波在可通行区域检测中的应用最重要的贡献是提供了一个算法融合的理论框架, 使得高效地综合利用边缘、纹理、色彩信息成为可能, 从而使移动机器人能够处理道路标线不理想、阴影、遮挡、多车道等城市环境中的常见问题. 卡尔曼系列的滤波器和粒子滤波都是贝叶斯理论框架下的滤波方

法, 在系统状态方程和测量方程非高斯或测量方程非线性的情况下, 都不能得到解析的最优贝叶斯解. 扩展卡尔曼滤波只能近似解决测量方程非线性的情况, 粒子滤波理论则通过重要性采样函数解决非线性和非高斯同时存在的问题<sup>[36, 37]</sup>. 因此, 重要性采样函数的选取是粒子滤波方法中的一个重要问题, 在现有的基于粒子滤波的结构化环境参数估计算法中<sup>[2, 14, 15]</sup>, 重要性采样函数多是基于 Bootstrap 算法<sup>[38]</sup>. 系统状态变量的选择通常与结构化道路下的选择类似, 比如  $x = (y_{\text{offset}}, b, c_0, c_1)^T$ , 其中 5 个参数分别代表移动机器人本体相对于环境的横向偏移、航向角偏差、可通行区域宽度、可通行区域横向和纵向曲率. 有时也简化为  $x = (y_{\text{offset}}, b)^T$ <sup>[2]</sup>.

设计粒子滤波器的另一个问题是似然函数  $p(z_k | x_k^{(i)})$  的选择, 用于在获取当前观测值后评估各种可能系统状态的权重, 其中  $x_k$  和  $z_k$  分别是  $k$  时刻的系统状态和观测值,  $i$  是粒子序号. 在可通行区域检测问题中,  $p(z_k | x_k^{(i)})$  可以没有解析形式, 通常表示如下<sup>[15]</sup>:

$$p_r(F_r | x) = \prod_{(x, y) \in \text{lane}} p_{\text{lane}}(F_r(x, y)) \times \prod_{(x, y) \notin \text{lane}} p_{\text{nonlane}}(F_r(x, y)). \quad (3)$$

其中  $F_r$  表示相应的特征图像(比如经过边缘检测后的图像),  $r = 1, 2, \dots, n$  代表当前使用的特征, 如色彩、边缘、纹理等. 在各特征相互独立的假设下, 如下框架便于多特征的融合:

$$p(F_1 \dots F_n | x) = \prod_{r=1, \dots, n} p_r(F_r | x). \quad (4)$$

$p_{\text{lane}}(F_r(x, y))$  的获取有多种方法, 最简单和常用的是用直方图来近似各特征的先验概率密度分布, 当前图像的特征直方图由上一帧结果统计得到. 这个动态更新的概率密度分布使得算法对光线和路面状况变化具有一定的抗干扰能力.

基于粒子滤波的道路检测在算法思想上与随机 Hough 变换或 RANSAC (RANDOM SAMPLE CONSENSUS) 算法相似, 都是“假设-验证”的过程. 相对于传统方法其主要改进体现在 3 个方面: 1) 环境模型的定义更加一般化, 不局限于有地面导引线的情形; 2) 提出了一个方法融合的框架, 高效率地综合利用边缘、纹理、色彩、灰度等不同特征; 3) 根据粒子的分布情况评估当前图像检测结果的可信度. 尤其是将各种传统的图像处理方法融合到一个理论框架下, 极大地提高了算法的鲁棒性. 实验表明, 这样的方法对地面阴影、光线变化有较强的适应能力, 通过

调整地面模型,还可处理相对复杂的情形,如多车道、无明显道路边界等。

虽然所使用的模型比结构化环境下的假设宽松了很多,但基于粒子滤波的方法仍然需要环境模型的假设,只是模型更加宽松.其中一个重要的改变就是道路宽度假设.在卡尔曼滤波系列的方法中,道路宽度被假定是正态分布的,当道路宽度发生较大变化时,滤波器可能发散.基于粒子滤波的方法通常采用的方法是假定道路宽度分布在一个可能的范围内,在检测过程中,粒子会收敛到正确的结果.但照顾到适应性不可避免会带来可靠性的问题,在多车道或不规则道路的情形下,且同时又有其他物体(比如其他移动物体)干扰时,很可能让滤波器不能正确收敛.此外,粒子滤波相对于传统滤波方法需要更大的计算量,因此必须根据具体的应用场合优化算法以满足实时需要.比如优先计算更可靠的特征,在计算资源紧张时自动放弃一些次要特征的分析.

作为独立的导航系统,基于粒子滤波的道路检测方法在理论上仍然需要更严密的可靠性分析,主要体现在初始化、意外情况的处理等方面.就目前的研究现状而言,该方法更适用于智能交通领域道路偏移警报系统,对于移动机器人导航只能作为辅助手段.

文献[39]介绍的系统可认为是这一阶段研究的最好代表,其结构可以图 2 进行概括.它为目的道路检测领域各个环节中使用最有效的方法提供了一个很好的框架.从图 2 中可以看出,基于粒子滤波理论的方法可应用于所有的步骤,实际上也正好对应了基于贝叶斯理论的滤波器的 3 类典型应用:特征评估、状态跟踪和系统辨识.

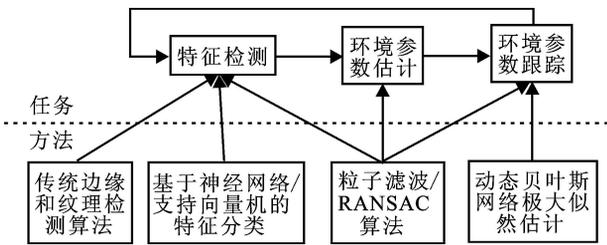


图 2 基于单目视觉的导航算法框架

### 5 针对非结构化环境的导航方法

越野环境是一个典型的非结构化环境.在该环境中,道路的概念变得更加模糊.基于二次曲线的道路模型将不再适用于结构化环境和非结构化环境,这时需要通过图像进行结构分析来定位移动机器人的“可通行区域”.在这类应用中,基于统计学习的方法是最广泛使用的工具.

### 5.1 早期针对越野环境的导航方法

基于统计学习理论的可通行区域检测方法最早可追溯到 SCARF 系统<sup>[40]</sup>,以及 CMU 早期的基于神经网络的系统 ALVINN<sup>[1]</sup>. SCARF 系统首先对地面色彩进行建模,生成一个地面色彩分布的高斯模型,用于给出每幅图像中所有像素属于路面和非路面的概率;然后将检测出的可通行区域与假定的环境模型进行匹配,最终找到一组最合适的道路参数.在设计环境模型时,考虑了道路分叉的情况,从而使得该方法具有一定的检测岔路口的能力. ALVINN 使用的是一个具有 4 个隐藏单元和 30 个输出单元的神经网络.所使用的训练输入是 30 × 32 分辨率的灰度图像,输出跳过了环境模型参数估计的阶段,直接给出的是移动机器人相对于环境的 30 个方向状态(已经包含了可能的障碍物信息),通过这些方向信息直接对机器人进行横向和纵向控制.该方法取得了较好的实验结果,但需要大量数据进行训练,而且以整幅道路图像作为系统输入的合理性缺少理论上的论证.

### 5.2 基于支持向量机的方法

文献[15]中,为了给出测量数据对当前粒子的评估,作者利用灰度直方图建立路面灰度的先验概率密度分布来得到似然函数  $p(z_k | x_k^{(p)})$ ,用于得到地面像素属于可通行区域的概率.建立可通行区域灰度直方图时,利用一个假设(即图像下方正中的位置为可通行区域,中部两侧的位置为非可通行区域),以建立最初级的灰度分布直方图.系统初始化时,这个步骤需要手动完成,由于缺少足够的样本,算法的可靠性可能存在隐患.文献[16]使用了一个基于支持向量机的方法来对像素进行分类(可同时给出像素属于可通行区域的概率),从而得到比上文提到的直方图方法更合理的效果.

文献[16]中使用的特征向量为  $x = \{R, G, B\}$ ,其中  $R, G, B$  分别为 RGB 色彩空间下的 3 个色彩分量.实际上,仅使用色彩空间的分量构成特征向量的方法仍不完善,像素的位置、纹理、色彩以及不同空间尺度(分辨率)下的特征包含有更丰富的信息,利用他们甚至可以重构出三维场景.所以,如何选取更有意义的特征是这类方法必须面临的一个重要问题. AdaBoost 算法可用来帮助寻找该特征<sup>[17,41]</sup>,它利用一些基本特征(如色彩、几何位置等)以及它们的线性、非线性组合生成高维的特征向量,通过组合弱分类器的方法进行分类.从相关实验可以看到,那些最终被选择的特征也可以以人类视觉的方式进行理解<sup>[41]</sup>.因此,通过这些经过分析的特征结合支持向量机的分类方法,将能得到更理想的结果.

类似的方法还有 RVM (Relevance Vector Machine)<sup>[42]</sup> 以及 IVM (Informative Vector Machine)<sup>[42]</sup>. RVM 方法通过在训练阶段控制支持向量(Support Vector)的数目来提高在线检测的速度. 由于训练通常可以离线进行, 通过 RVM 方法裁剪支持向量, 再以 IVM 方法对路面像素块进行分类, 可以在一定程度上满足实时应用的需要, 检测效率可达到 5 帧/s. 但这些方法的优劣并不容易比较, 最终的性能除了算法本身, 还取决于算法的具体实现, 甚至代码的质量, 比如训练的方法、特征的选取等.

### 5.3 基于单幅图像的三维重构

如果能从单幅图像重构出移动机器人周围的三维场景, 则基于单目视觉的导航将进入一个全新的阶段. 因为在实际应用中, 大多数室外移动机器人都需要面临不确定的非结构化环境. 但平面图像丢失了深度信息, 所以从二维图像的局部特征无法直接重构三维场景. 立体视觉方法是进行三维重构的一个可行方案, 但立体视觉面临 3 个问题: 视场有限、设备昂贵、室外环境下算法的效率和可靠性很难保证. 由于这些局限, 近年来, 研究者开始在单目视觉方法上进行三维场景重构的探索<sup>[19,20]</sup>, 其基本思路是通过分析不同空间尺度(分辨率)上像素(像素块)之间关系, 通过机器学习方法来获得丢失的深度信息. 对于移动机器人导航而言, 这个研究的意义在于可以借此定位出“可通行区域”, 得到一个类似于距离传感器(如激光雷达、超声波传感器等)生成的栅格地图的局部地图信息. 文献[19]提出了一种区分路面、建筑、天空等区域的方法, 而[20]中更是直接重构出了场景的三维模型, 实际上已可以提供完整的局部导航信息. [20]中用于获取场景绝对深度的信息包括 9 个  $3 \times 3$  Laws' Masks 算子、6 个纹理梯度算子和 2 个 YCbCr 色彩空间下的色彩通道的均值滤波, 共 17 个特征, 以不同分辨率下的“像素块”(patch)为单位进行处理, 根据实验选取每个像素块经过这些算子处理后各像素值的绝对值之和以及平方和生成特征向量. 这样初始的特征向量为 34 维 ( $17 \times 2$ ). 除像素块本身外, 还考虑了 3 个不同尺度(分辨率)下的 4 邻域像素块的特征, 以及当前像素块所在图像列的 4 个列像素块的特征, 这样每个像素块对应的特征向量的维数将达到  $34 \times (5 \times 3 + 4) = 646$  维. 图 3 即为[20]中介绍的导航系统. 其训练样本中的真实深度图像由一个伪三维激光雷达(带一个自由度云台的平面激光雷达)生成. 目前这个算法分析一幅图像可能需要几分钟甚至几十分钟时间, 还不能满足实时导航的要求, 特征选取也有改进

余地. 但从实际应用的角度看, 仍可用来初始化导航系统, 部分解决移动机器人的“绑架”问题. 即通过环境类型辨识选取合适的环境模型, 当对环境有了初步理解以后, 再使用针对特定环境的经典方法进行导航. 对于这类方法, 随着理论的成熟、算法的改进和硬件水平的提高, 处理一幅图像所需要的时间将自然减少. 所以从应用的角度看, 提高这类方法结果可靠性的努力比提高算法效率更迫切.



(a) 真实场景 (b) 激光雷达 (c) 单目重构 (d) 实验平台

图 3 基于单目视觉的三维重构

## 6 结论

面向室外移动机器人的视觉导航研究将分化为两个方向: 传感器融合技术的发展与基于单目视觉的三维重构方法的完善. 由于后者的算法可靠性和效率仍有待进一步论证, 视觉方法与惯性导航元件以及距离传感器的融合在可以预见的未来仍然是实用系统的主要研究方向. 单目视觉在移动机器人多传感器导航系统中将主要承担环境模型辨识、结构化环境参数估计、道路偏移检测或特殊道路标志的检测和识别等工作. 在城市环境道路这样的半结构化环境下, 视觉导航系统较倾向于通过方法融合来提高算法的可靠性. 而在排爆、消防机器人所面临的完全非结构化环境中, 则只能依靠基于单幅图像的三维重构技术才能开发出实际可用的导航系统. 这个领域的研究目前虽然与实时应用还有距离, 但它生理理论上可以部分解决机器人的绑架问题, 随着相应的数学方法的改进和硬件水平的提高, 将会有很大的发展空间. 即使在目前不能满足实时导航应用的情况下, 仍可作为系统初始化的工具和环境模型辨识的参考. 它与 GPS 的结合使用, 将是基于单目视觉的室外移动机器人导航系统完全走出场景假设的第一步.

### 参考文献(References)

- [1] Pomerleau D. Neural network vision for robot driving [C]. In the Handbook of Brain Theory and Neural Networks. Cambridge: MIT Press, 1995: 161-181.
- [2] Apostoloff N. Vision based lane tracking using multiple cues and particle filtering[D]. Canberra: The Australian National University, 2005.
- [3] Se S, Lowe D, Little J. Mobile robot localization and mapping with uncertainty using scale-invariant visual

- landmarks[J]. *The Int J of Robotics Research*, 2007, 14(2): 735-758.
- [4] Lutzeler M, Dickmanns E D. Road recognition with MarVEye [C]. *IEEE Int Conf Intelligent Vehicles*. Stuttgart, 1998: 341-346.
- [5] Bertozzi M, Broggi A. GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection[J]. *IEEE Trans on Image Processing*, 1998, 7(1): 62-81.
- [6] Pomerleau D, Jochem T. Rapidly adapting machine vision for automated vehicle steering[J]. *IEEE Expert: Special Issue on Intelligent System and their Applications*, 1996, 11(2): 19-27.
- [7] Kreucher C, Lakshmanan S. LANA: A lane extraction algorithm that uses frequency domain features[J]. *IEEE Trans on Robotics and Automation*, 1999, 15(2): 343-350.
- [8] Gregor R, Lutzeler M, Siedersberger K H, et al. EMS-vision: A perceptual system for autonomous vehicles [J]. *IEEE Trans on Intelligent Transportation Systems*, 2002, 3(1): 48-59.
- [9] Labayrade R, Douret J, Aubert D. A multi-model lane detector that handles road singularities[C]. *Proc of the IEEE Intelligent Transportation Systems Conf*. Toronto, 2006: 1143-1148.
- [10] Davies B, Lienhart R. Using CART to segment road images [C]. *SPIE Multimedia Content Analysis, Management, and Retrieval*. San Jose, 2006: 60730U01-60730U12.
- [11] Hummel B, Kammel S, Thao D, et al. Vision-based path-planning in unstructured environments[C]. *Proc of the IEEE Intelligent Vehicle Symposium*. Tokyo, 2006: 176-181.
- [12] Rasmussen C. Texture-based vanishing point voting for road shape estimation [C]. *British Machine Vision Conf*. Kingston, 2004: 47-56.
- [13] Lombardi P, Zanin M, Messelodi S. Switching models for vision-based on-board road detection[C]. *Proc of the IEEE Conf on Intelligent Transportation Systems*. Vienna, 2005: 67-72.
- [14] Smuda P, Schweiger R, Neumann H, et al. Multiple cue data fusion with particle filters for road course detection in vision systems [C]. *Proc of the IEEE Intelligent Vehicle Symposium*. Tokyo, 2006: 400-405.
- [15] Franke U, Loose H, Knoppel C. Lane recognition on country roads [C]. *Proc of the IEEE Intelligent Vehicles Symposium*. Istanbul, 2007: 99-104.
- [16] Zhang H, Hou D, Zhou Z. A novel lane detection algorithm based on support vector machine [C]. *Progress in Electromagnetics Research Symposium*. Hangzhou, 2005: 390-394.
- [17] Sha Y, Zhang G Y, Yang Y. A road detection algorithm by boosting using feature combination [C]. *Proc of the IEEE Intelligent Vehicles Symposium*. Istanbul, 2007: 364-368.
- [18] Hoiem D, Efros A A, Hebert M. Geometric context from a single image [C]. *ICCV 2005*. Pittsburgh, 2005: 654-661.
- [19] Hoiem D. Seeing the world behind the image - spatial layout for 3D scene understanding [D]. Pittsburgh: Carnegie Mellon University, 2007.
- [20] Saxena A, Chung S H, Ng A Y. 3-D depth reconstruction from a single still image [J]. *Int J of Computer Vision*, 2008, 76(1): 53-69.
- [21] Dickmanns E D, Mysliwetz B D. Recursive 3-D road and relative ego-state recognition [J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1992, 14(2): 199-213.
- [22] Jung C R, Kelber C R. A lane departure warning system based on a linear-parabolic lane model [C]. *Proc of the IEEE Intelligent Vehicles Symposium*. Parma, 2004: 891-895.
- [23] Donald J B M, Franz J, Shorten R. Application of the hough transform to lane detection in motorway driving scenarios [C]. *Proc of the Irish Signals and Systems Conf*. Maynooth, 2001: 340-345.
- [24] Schreiber D, Ales B, Clabian M. Single camera lane detection and tracking [C]. *IEEE Conf on Intelligent Transportation Systems*. Vienna, 2005: 302-307.
- [25] Collado J M, Hilario C de la Escalera A, Armingol J M. Detection and classification of road lanes with a frequency analysis [C]. *Proc of the IEEE Intelligent Vehicles Symposium*. Las Vegas, 2005: 78-83.
- [26] Aufrere R, Chapuis R, Chausse F. A model-driven approach for real-time road recognition [J]. *Machine Vision and Application*, 2001, 13(2): 95-107.
- [27] Bertozzi M, Broggi A, Fascioli A, et al. An evolutionary approach to lane markings detection in road environments [C]. *Proc of the Congress of the Italian Association for Artificial Intelligence*. Siena, 2002: 627-636.
- [28] Boumediene M, Ouamri A, Dahnoun N. Lane boundary detection and tracking using NNF and HMM approaches [C]. *Proc of the IEEE Intelligent Vehicles Symposium*. Istanbul, 2007: 1107-1111.
- [29] Sehestedt S, Kodagoda S, Alempijevic A, et al. Efficient lane detection and tracking in urban environments [C]. *European Conf on Mobile Robots*. Freiburg, 2007: 1-6.

(下转第 19 页)

- Pattern Recognition. Berlin: Springer-Verlag, 2004: 662-670.
- [9] Wagstaff K, Cardie C, Rogers S, et al. Constrained  $K$ -means clustering with background knowledge[C]. Proc of 18th Int Conf on Machine Learning. Williamstown: Morgan Kaufmann Publishers, 2001: 577-584.
- [10] Basu S, Banerjee A, Mooney R. Active semi-supervision for pairwise constrained clustering [C]. Proc of the SIAM Int Conf on Data Mining. Cambridge: MIT Press, 2004: 333-344.
- [11] Grira N, Crucianu, Boujemaa N. Semi-supervised fuzzy clustering with pairwise-constrained competitive agglomeration[C]. IEEE Int Conf on Fuzzy Systems. Reno: IEEE Press, 2005: 22-25.
- [12] Davidson I, Ravi S S. Agglomerative hierarchical clustering with constraints: Theoretical and empirical results [C]. Proc of the 9th European Conf on Principles and Practice of Knowledge Discovery in Databases. Porto: Springer, 2005: 59-70.
- [13] Davidson I, Ravi S S. Clustering with constraints: Feasibility issues and the  $K$ -means algorithm[C]. Proc of the 5th SIAM Int Conf on Data Mining. Newport Beach, 2005: 138-149.
- [14] Davidson I, Ravi S S. Intractability and clustering with constraints [C]. Proc of 24th Int Conf on Machine Learning. New York: ACM Press, 2007: 201-208.
- [15] Wei T, Hui X, Shi Z, et al. Enhancing semi-supervised clustering: A feature projection perspective [C]. Proc of the 13th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM Press, 2007: 707-716.
- [16] 张西芝, 姬波, 邱保志. 基于网格的多密度聚类算法[J]. 微计算机信息(管控一体化), 2005, 21(36): 101-103.  
(Zhang X Z, Ji B, Qiu B Z. Grid-based clustering algorithm for multi-density [J]. Control & Automation, 2005, 21(36): 101-103.)

---

(上接第 7 页)

- [30] Xu L, OJA E. Randomized hough transform (RHT): Basic mechanisms, algorithms and computational complexities[J]. Computer Vision Graphics and Image Processing, 1993, 57(2): 131-154.
- [31] Li Q, Zheng N, Cheng H. Lane boundary detection using an adaptive randomized hough transform [C]. Proc of the 5th World Congress on Intelligent Control and Automation. Hangzhou, 2004: 4084-4088.
- [32] Li Q, Zheng N, Cheng H. Springrobot: A prototype autonomous vehicle and its algorithms for lane detection [J]. IEEE Trans on Intelligent Transportation Systems, 2004, 5(4): 300-308.
- [33] Yu B, Jain A. Lane boundary detection using a multi-resolution hough transform [C]. Int Conf on Image Proc. Washington D C, 1997: 748-751.
- [34] Wang Y, Teoh E K, Shen D. Lane detection using B-snake [C]. Proc of Information Intelligent Systems. Bethesda, 1999: 438-443.
- [35] Wang Y, Teoh E K, Shen D. Lane detection and tracking using B-snake [J]. Image and Vision Computing, 2004, 22(4): 269-280.
- [36] Arulampalam S, Maskell S, Gordon N, et al. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking[J]. IEEE Trans on Signal Processing, 2001, 50(2): 174-188.
- [37] Doucet A, Godsill S, Andrien C. On sequential monte carlo sampling methods for Bayesian filtering [J]. Statistic and Computing, 2000, 10(3): 197-208.
- [38] Gordon N J, Salmond D J, Smith A F M. Novel approach to nonlinear/non-Gaussian Bayesian state estimation[C]. IEE Proc-F, 1993, 140(2): 107-113.
- [39] Kim Z W. Robust lane detection and tracking in challenging scenarios [J]. IEEE Trans on Intelligent Transportation Systems, 2008, 9(1): 16-26.
- [40] Crisman J D, Thorpe C E. SCARF: A color vision system that tracks roads and intersections [J]. IEEE Trans on Robotics and Automation, 1993, 9(1): 49-58.
- [41] Bahlmann C, Zhu Y, Visvanathan Ramesh Pellkofer M, et al. A system for traffic sign detection tracking and recognition using color shape and motion information[C]. Proc of the IEEE Intelligent Vehicles Symposium. Las Vegas, 2005: 255-260.
- [42] Tipping M E. Sparse Bayesian learning and the relevance vector machine [J]. J of Machine Learning Research, 2001, 1(1): 211-244.
- [43] Lawrence N, Seeger M, Herbrich R. Fast sparse Gaussian process methods: The informative vector machine [C]. Advances in Neural Information Processing Systems. Cambridge, 2003: 609-616.