

# 移动机器人基于视觉室外自然场景理解的研究与进展

庄严<sup>1,2</sup> 陈东<sup>1,2</sup> 王伟<sup>1</sup> 韩建达<sup>2</sup> 王越超<sup>2</sup>

**摘要** 对于工作在典型非结构化场景中的移动机器人系统,具有良好的室外自然场景感知与理解能力是其能够自主运行的前提条件。移动机器人使用视觉传感器来进行室外自然场景的理解一直是该领域的研究热点。本文首先介绍了基于视觉的移动机器人自然场景理解的研究现状,对其相关子领域的研究思路与前沿技术进行了着重论述与分析,并从实时性和环境自适应性等方面对相关技术的实用性问题加以讨论。最后对该领域的研究重点和技术发展趋势进行了探讨。

**关键词** 自然场景理解, 自主移动机器人, 场景表达, 统计建模, 计算机视觉, 认知学习  
**中图分类号** TP24

## STATUS AND DEVELOPMENT OF NATURAL SCENE UNDERSTANDING FOR VISION-BASED OUTDOOR MOBILE ROBOT

ZHUANG Yan<sup>1,2</sup> CHEN Dong<sup>1,2</sup> WANG Wei<sup>1</sup> HAN Jian-Da<sup>2</sup> WANG Yue-Chao<sup>2</sup>

**Abstract** Natural scene understanding is an indispensable capability for vision-based outdoor mobile robots working in the challenging unstructured environment. The intention of this paper is to investigate the status and development of this technology, point out and analyze some cutting-edge technologies in the relevant sub-domains. Practical issues are discussed throughout the paper in terms of real-time and self-adaptation aspects. The research emphases and prospective technical development trends are also proposed at the end of this paper.

**Key words** natural scene understanding, autonomous mobile robot, scene representation, statistical modeling, computer vision, cognitive learning

### 1 引言 (Introduction)

如何使机器人能更好的理解其所工作环境,或具有与智能生命体相类似的环境认知能力,是长久以来国内外学者密切关注并积极研讨的具有挑战性的研究课题之一。在相对结构化的室内环境中,借助于多传感器融合技术的移动机器人自主环境感知、环境地图构建以及室内场景认知技术相对成熟<sup>[1]</sup>。因此近年来自主移动机器人(包括自主无人驾驶车)的研究与应用正逐步从室内结构化环境向野外完全非结构化环境进行扩展。基于视觉的室外自然场景理解是工作在复杂自然环境中的移动机器人能够实现自主环境适应所应具备的基本条件。由于室外自然场景的多样性、随机性、复杂性以及移动机器人的运动性,所构建的场景理解系统应具有较高的实

时性和自适应性。实时性是指由于自然场景图像本身的不稳定性和复杂性,为了提高辨识效果往往会造成图像处理的时间开销过大,因此必须要兼顾算法的效率与辨识效果。同时为了应对所处环境中自然景物的非结构化特性和随机性,以及景物在不同地形地貌之中的相互组合与关联,算法的自适应性也是决定自然场景理解效果的重要因素。

目前视觉场景理解还没有严格统一的定义。参考 MIT、CMU 和 Stanford 等大学的国际著名科研团队的研究工作<sup>[2-4]</sup>,视觉场景理解可表述为在环境数据感知的基础上,结合视觉分析与图像处理识别等技术手段,从计算统计、行为认知以及语义等不同角度挖掘视觉数据中的特征与模式,从而实现场景有效分析、认知与表达。近年来结合数据学习与挖掘、生物认知特征和统计建模方法构建的视觉场景认知理解系统,为室外场景辨识和物体识别提供了新的解决方案。其中最具代表性的是由 DARPA 主办的野外无人车挑战赛<sup>[5]</sup>,参赛无人车在室外复杂场景下的深层环境感知和稳定运行推动了自然场景理解在实际平台上的技术转化。但由于室外自然场景的多样性与复杂性,传统室内移动机器人的辨识与认知技术无法轻易的加以转移利用,这就为室外自然场景理解提出了新的研究需求与技术挑战。

收稿日期 XXXX-XX-XX 收修改稿日期 XXXX-XX-XX  
Received Month Date, Year; in revised form Month Date, Year  
此项工作得到国家自然科学基金(60605023,60775048),国家863项目(2007AA04Z257)和机器人学国家重点实验室开放课题(RLO200802)资助  
Supported by National Natural Science Foundation of China(60605023,60775048),National High Technology Research and Development Program of China(2007AA04Z257) and State Key Laboratory of Robotics(RLO200802)  
1. 大连理工大学信息与控制研究中心 大连 116024 2. 机器人学国家重点实验室 沈阳 110016  
1. Research Center of Information and Control, Dalian University of Technology, Dalian 116024 2. State Key Laboratory of Robotics, Shenyang 110016  
DOI: 10.3724/SP.J.1004.2008.xxxxx

## 2 研究现状 (Literature review)

早期的相关工作主要集中于物体的分割辨识及认知基础理论<sup>[6,7]</sup>的研究上, Fischler<sup>[8]</sup>给出物体匹配准则, 但其局限于物体的几何特征上。Biederman<sup>[9]</sup>提出组件辨识理论 (RBC) 来对人类图像理解过程进行建模, 但是该理论只是局限在具有确定边界的特定实体的辨识上。此时物体辨识所处的场景相对简单, 对物体和场景本身施加的约束较多, 但其对物体辨识过程中遮挡、光照、变形等影响因素的思考和析为后来室外自然场景的理解奠定了基础。

90年代 Pomerleau<sup>[10]</sup>、Jochem<sup>[11]</sup>利用离线训练得到的人工神经网络模型完成了室外道路环境下较低层次的理解。由于处理的数据量相对较少, 同时采用监督学习策略, 算法的实时性高, 但并没有实现对室外复杂自然场景的自适应理解。Walter<sup>[12]</sup>提出条件规则产生 (CRG) 方法, 该方法通过建立分类规则可对复杂场景中多类物体进行辨识, 其思想对室外自然场景理解有一定启发性。Draper<sup>[13]</sup>根据当时技术发展状况指出无法完成室外复杂场景理解的两点主要原因: 第一是视觉算法的不成熟, 但随着各子领域技术的进步该问题已有很大改进; 第二是缺少对视觉算法流程的控制策略, 这一点是阻碍该研究发展的关键。同时作者基于对室外景物多变性、模糊性、表征的多层次性等问题的分析理解, 指出集成各子领域的研究成果来构建室外自然场景理解系统的时机已经相对成熟。Battle<sup>[14]</sup>针对上述不足分析了当时运用于室外自然场景感知与理解的视觉系统, 指出基于混合结构的视觉算法可以较好的适应室外复杂场景。但是受限于当时计算机处理速度等技术因素, 所开发的视觉系统大多不能满足室外移动机器人平台对实时性或自适应性的要求。

2000年以来, 随着计算机视觉技术和认知学的快速发展, 对场景图像中单一形式物体的识别逐渐过渡到对类内多形式物体的识别<sup>[15]</sup>, 以及场景的全局理解<sup>[2,16]</sup>和场景物体间关联信息的建模<sup>[17]</sup>。从场景建模角度来看, 马尔可夫随机场 (MRF)<sup>[18]</sup>、条件随机场 (CRF)<sup>[19]</sup>等模型的使用增强了场景信息的关联, 这使得场景及场景中物体辨识的速度和精度都有所提高。从场景特征描述的角度来看, Lowe<sup>[20]</sup>给出场景尺度不变描述符 SIFT, 该局部描述符对尺度、平移、旋转等变换具有较好的不变性。Torralla<sup>[21]</sup>提出共享特征机制并通过多任务学习获得不同物体的公共特征, 从而减少了多物体识别的计算量。Shotton<sup>[22]</sup>提出了 texton-layerout 高级特征并结合 boosting 方法, 有效解决了室外环境中辨识模糊性问题, 使场景及场景中物体的辨

识具有较高的精度, 但算法效率不高且只是孤立地考虑每帧图像的信息。Fauqueur<sup>[23]</sup>充分利用视频帧间的关联信息, 采用局部 SIFT 特征进行帧间区域跟踪和标记传播, 降低了场景辨识的计算量, 同时该工作为采用视频分析手段进行自然场景理解开拓了思路。由于不同于结构化建筑物和室外道路这些特定场景<sup>[24,25]</sup>, 自然场景具有非结构化特性和随机性。由于光照、地形及运动等多种因素会使得自然景物在视觉系统中的成像有显著差异, 这增加了对非规则自然景物识别与理解的难度。现今的研究大多集中于自适应地对场景进行快速理解, 并利用多传感器信息融合技术提高自然场景理解的鲁棒性。在实时性方面, Wang<sup>[26]</sup>利用视觉局部显著区域来对自然场景进行辨识和理解, 通过提取场景认知过程中的感兴趣区域, 大大提高了算法的实时性。Huhns<sup>[27]</sup>通过训练获得不同场景类别颜色分量阈值, 进而采用基于规则的方法对不同地貌进行分类辨识, 所构建的系统具有良好的实时性。在多传感器信息融合方面, Kelly<sup>[28]</sup>从不同种类视觉传感器获得场景信号的多角度表征, 进而对场景从不同方面加以测量, 从而有效地对不同地形进行检测和分类。Schafer<sup>[29]</sup>利用激光、立体视觉传感器获得场景的深度信息, 进而检测岩石、树木、水坑等非连续地形信息。在自适应性方面, Thrun等<sup>[30]</sup>利用自监督学习方法, 克服了有监督学习环境适应能力差和半监督学习算法不稳定的缺点, 成功对大范围场景进行预测辨识。

由于该研究领域的多学科交叉特点, 其所取得的研究成果要得益于图像处理与分析、模式识别与分类、机器学习、知识工程等子领域的研究进展, 同时各子领域间的相互交融也极大的推动了该领域研究的创新与发展。本文尝试从自然场景理解研究的实时性、自适应性、稳定性和智能性等方面对该领域进行分析与探讨, 其中各个方面的研究可相互补充与完善, 包括以图像处理为主的多层次场景图像表达, 基于不同学习策略的场景信息有效学习, 确保视觉系统鲁棒性的系统知识结构与视觉控制策略, 以及与生物认知学相结合的自然场景理解。最后给出了该领域的研究重点和技术发展趋势。

## 3 多层次场景图像表达

如果将自然场景图像理解按照低、中、高三种表达层次进行划分, 可对应分为自然场景图像分割、图像区域标记和图像分类与注释三个层次。其中作为中间层的图像区域标记是利用图像低层的颜色、纹理等特征或高层的全局场景特征来形成图像不同区域的语义信息。该信息是连接低层和高层的桥梁, 可克服高低两层之间表达跨度较大所产生的弊端,

因而更加具有挑战性。以下分别从这三个层次加以阐述。

### 3.1 自然场景图像区域分割

Zhang<sup>[31]</sup> 回顾了过去 40 年的图像与视频分割技术,并将分割算法分为四类:并行边界技术、串行边界技术、并行区域技术和串行区域技术。对于串行技术,如区域的生长、分裂、合并以及边界的闭合等,后期的处理往往要利用到前期的处理结果。而并行技术在处理过程中所有的判断和决定都可独立和同时地做出,所以具有较高的计算效率。为满足室外场景理解的实时性要求通常选用并行分割技术。

自然场景分割比较典型的算法包括:基于边界的分割<sup>[32]</sup>、基于特征的 Mean-Shift 方法<sup>[33]</sup>、基于图论的 NCuts 方法<sup>[34]</sup>、基于能量函数优化的分割<sup>[35]</sup>等。文献<sup>[36]</sup>从统计角度给出了自然场景分割具有模糊性的两点原因:(1)在相同的尺度空间中,不同自然图像的局部特征(如颜色、边界、纹理等)的统计特性不具有均匀性。即使是同一幅图片的不同区域也存在该问题,所以应采用多层次多尺度的分割方法。(2)由于图像不同区域纹理特征的复杂性,即使考虑了尺度因素也不易确定合适的特征维数。因此文献<sup>[36]</sup>假设自然图像的纹理特征服从混合高斯分布,并且遵循最优的分割应使图像特征编码长度最短这一准则。

从近年来该领域研究的发展趋势可以看出,将场景分割与分类辨识两个过程区分开来的研究方法有相当的局限性,这是因为单独将场景分割成没有形成语义信息的不同区域缺乏实用意义,所以将分割和区域分类相融合成为当前研究的主要趋势。同时为了保证场景理解的实时性,图像分割所占的比重相对弱化,而面向特定应用的分割方法则被广泛采用,即将场景中移动机器人感兴趣的目标分割出来,这主要包括可行区域和障碍区域两类。这样有针对性地对上述目标进行分割辨识可显著地降低计算量,从而满足实时性的要求。

### 3.2 自然场景图像区域标记

作为自然场景表达的中间层次,区域标记技术最终要对场景图像的各不同区域加以解释,由于场景的复杂性使得辨识的难度加大,计算量也随之上升。所以辨识过程要同时兼顾辨识效果与速度。PASCAL<sup>[37]</sup>物体识别挑战赛的举办推动了这方面的发展。

Shotton<sup>[22]</sup>选取图像局部区域内的 texton 个数作为特征,采用 boosting 算法训练得到大量置信度较低的弱分类器,进而联合得到可取得较好分类效果的强分类器,从而达到对纹理、形状和场景语境信息的联合建模。由于算法需要经过多次滤波得到

texton 特征以及 K 均值聚类的耗时性造成了算法的低效率,并且需要占用大量内存空间,但该方法所提出的 TextonBoost 分类辨识思想对该领域的研究工作具有良好的启发性。在随后研究中,Shotton<sup>[38]</sup>通过训练得到多个随机决策树并将其构建成随机决策森林以用于快速分类辨识。类似于文本分析中的 bag-of-words 方法,每个决策树的分支节点和叶子节点分别对应各自的物体类别分布直方图。由于森林中每个决策树所生成的直方图相当于弱分类器,因此综合多决策树直方图便可得到较好的分类效果。随机决策树的使用免去了 TextonBoost 方法中耗时的滤波和聚类,因此算法具有高的实时性。

Verbeek<sup>[39]</sup>综合了基于概率的潜在语义分析模型(pLSA)和 MRF 的优点来对图像区域进行分类辨识。潜在语义分析模型仅对局部区域进行分类辨识,却没有考虑图像空间结构的关联信息,而马尔可夫随机场模型可对场景的空间信息形成约束,因而将二者结合能够取得实时的辨识效果。在文献<sup>[40]</sup>中,Verbeek 通过训练进一步从部分标记的场景图像中学习得到条件随机场模型,该模型包含场景空间局部约束信息和场景的全局语境信息。由于训练图像集只需进行部分标记,因而降低了大量耗时的手工标记工作,单幅测试图像辨识时间在 3 秒左右。

### 3.3 自然场景图像分类与注释

图像分类技术对应于图像的高层表达,是对整幅图像信息的解释,也称为图像注释。文献<sup>[41]</sup>给出图像分类的层次结构。首先将图像区分为室内环境和室外环境,然后进一步将室外环境区分为城市环境和自然环境,而对自然环境又可细分为河流、森林、山地等不同环境。对于每一层次采用不同的场景特征与之相对应,通过建立混合高斯模型并依据最大后验概率准则得到每一层次的分类结果。

从语义角度<sup>[42]</sup>可将其大致分为对象语义方法、局部语义方法和全局语义方法。

(1)对象语义方法:首先通过图像区域标记技术得到场景中的景物信息,然后根据景物的分布情况给出场景分类结果。文献<sup>[43]</sup>利用已标记的场景图像通过训练得到一组检测函数,每个检测函数用于检测不同类别物体。对场景中景物类别的分布信息进行建模,并使用 EM 算法确定最优模型参数,再依据最大后验概率准则得到测试图片的分类结果。

(2)局部语义方法:借鉴于文本分析中的 bag-of-words 方法,将场景比作文本,将场景中存在的景物比作文本主题,不同的景物对应不同的主题。进行辨识的场景往往是多主题的混合,从而主题的分布信息(视觉词汇直方图)决定了场景的分类结果。文献<sup>[44]</sup>采用 LDA(Latent Dirichlet Analysis)模型,

文献<sup>[45]</sup>采用 pLSA(probabilistic Latent Semantic Analysis)模型,二者将图片分成固定大小的网格,每个网格使用归一化的灰度值和 SIFT<sup>[20]</sup>描述符作为特征向量,通过学习生成不同场景的词汇表,对场景具有较高的识别率。由于局部语义方法时间开销大,通常不满足移动机器人实时性的要求。

(3) 全局语义方法:该方法通过提取场景的全局统计信息从而达到对场景的整体理解。文献<sup>[2,46]</sup>给出了认知场景的五种全局特征:自然度(naturalness),开放度(openness),粗糙度(roughness),扩展度(expansion),不平整度(ruggedness)。利用这五种特征形成了新的特征空间(spatial envelop)。该方法不必考虑场景的局部特征和场景中存在的景物信息,仅从全局信息出发即可将场景划分为山地、森林、沙滩等 6 种类别。由于往往与自顶向下的算法结构相结合,所以实际自然场景辨识过程中具有较高的效率。

图像分类技术所得到的结果往往作为先验信息来辅助自然场景的理解,这类似于 5.1 节中的知识结构,即根据某一场景分类信息进一步预测场景中可能存在的景物<sup>[47]</sup>,从而提高理解的速度和精度。

#### 4 场景信息的有效学习

与早期机械地执行一组指令集的单一任务系统相比,工作在室外场景中的自主移动机器人系统必须要有良好的自学习能力。现有的基于有监督学习的自主系统<sup>[48]</sup>需要在离线状态下对手工标记后的数据进行训练,而基于视觉来评估地形可通过能力的机器人系统<sup>[49]</sup>要假设可通过区域的特性相对固定。由于只能工作在预先设定的室外环境中,而无法适应超出训练集以外的环境改变,因此二者都有较大的应用局限性。

为了推动该领域研究的发展,DARPA 最近 3 年的 LAGR(Learning Applied to Ground Robots)研究项目共资助 10 个相关的研究组来进行室外不规则环境下移动机器人自导航软件的开发。研发项目要求将学习策略与视觉算法相融合,使移动机器人能对未知环境中的地形起伏、光照强弱变化、天气季节影响等因素均具有较好的自主适应能力。各研究组采用自监督学习、半监督学习和主动学习等方法取得初步研究成果。其中基于主动学习<sup>[50]</sup>和半监督学习<sup>[51]</sup>的系统不但避免了有监督学习需要对场景图片进行手工标记的局限,而且对复杂场景的在线辨识效果要优于完全有监督的系统。然而现有的半监督学习方法性能通常不太稳定,如何有效改善学习性能仍有待进一步解决。

无人车 RoboJeep<sup>[52]</sup>对环境状态信息、单幅图像的颜色分布信息和已标记的地貌颜色分布信息建

立三种不同类型的混合高斯分布模型,对多自然景物具有高的辨识率。由于实际中多种因素的影响,不可能用单一模型或物质原型来代表同类物体,基于样例的方法模拟人的学习过程,认为对同类物体的认识是与记忆中已存在的该物体典型样例的相似度比较获得的。文献<sup>[53]</sup>采用该方法,通过在线更新样例库,取得了良好的辨识效果。文献<sup>[54]</sup>结合了有监督和无监督学习方法来评估场景的可通过能力。其中离线阶段采用神经网络训练得到场景几何特征与可通过性之间的关系,而在线阶段通过学习得到场景的颜色分布和几何特征间的关联,这样即便在缺少场景几何信息时仍可通过场景的颜色模型进行可通过性预测。为了更好地适应各种复杂因素,近年来开发的系统较多采用自监督学习方法,该方法不但可以减少或省去对场景图片的手工标记,而且在线辨识效果受环境变化的影响较小。该方法首先通过较可靠的传感器(如激光雷达、立体视觉等)获得近处可通过区域的数据,并利用该数据进行在线训练从而实现了大范围场景的辨识,所以也称为由近到远的学习方法。文献<sup>[55]</sup>利用机器人获取的卫星图像和激光测距数据进行在线训练,从而得到工作场景的概率模型,并用于环境可通过性的预测。文献<sup>[56]</sup>采用自监督学习对地面粗糙度进行预测。Stanley<sup>[57]</sup>无人车系统首先采用激光对近处地面进行可通过性估计,之后对该区域进行采样,得到的数据作为训练样本,依据颜色直方图建立概率模型,进而利用该模型对远景进行分析预测。Hadsell<sup>[58]</sup>采用自监督学习策略,通过训练得到的分级网络来提取每帧图像的有效特征,进一步利用该特征训练得到实时的长距离分类器,从而实现大范围场景可通过性判断,系统对未知环境具有快速的适应能力。以上学习方法大多只停留在对可通过区域或未知障碍区域的检测和识别,如何实现自然场景的自适应理解,仍然是一个具有挑战性的课题。

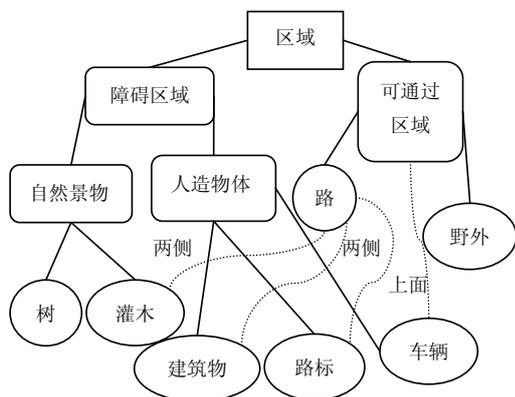
#### 5 系统知识与视觉算法控制策略

为了保证移动机器人进行室外自然场景理解的稳定性和自适应性,系统除了要具有实时的图像处理和分析能力以及优良的学习算法外,还应具有完善的系统知识结构和视觉控制策略。虽然不少学者从系统工程和知识工程学的角度来理解、分析并构建基于不同结构的系统知识库,但对视觉控制策略的研究则相对有限。

##### 5.1 系统知识结构

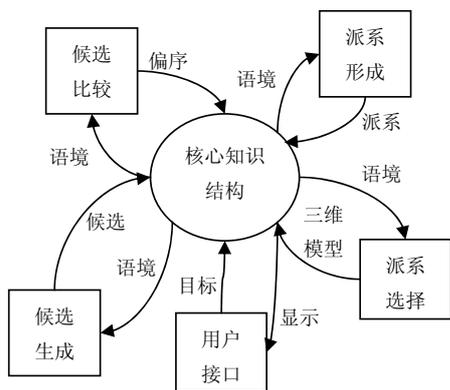
由于各种不确定因素的存在,对于室外自然场景中的景物辨识具有一定的模糊性,并经常导

致分类的错误。比如水面和天空的纹理特征较为相似,很可能辨识的结果会将水误认为天空。为解决该类问题就需要对场景加以约束,即构建场景的知识库,其包括景物信息以及景物间的关联相互关系信息。早期的视觉系统,如 VISION Schema<sup>[59]</sup>、CONDOR<sup>[60]</sup> 等,采用知识工程学的手段来组织场景信息,这类显式表达的信息是对场景中绝对信息的表述。文献<sup>[14,61]</sup> 分析总结了典型的知识组织结构,包括方案(Schema)、黑板(Blackboard)、关系图、网络等(如图 1 所示)。对于构建通用的机器人室外认知系统,具有易操作性、易扩展性和易更新性的知识结构是必不可少的。此外文献<sup>[61]</sup> 还对其它种类的知识进行分类与概述,其中包括图像处理方法的选择、基于特定任务的知识和用户接口知识等,这些都是整体知识框架的辅助信息。



(a) 关系图知识结构

(a) relation graph knowledge structure



(b)CONDOR 系统知识结构

(b)knowledge structure of CONDOR system

图 1 关系图知识结构和 CONDOR 系统知识结构

Fig. 1 relation graph knowledge structure and knowledge structure of CONDOR system

随着计算机视觉算法的快速发展,对场景信息的表达较多的采用基于模型的手段,如 HMM、MRF、CRF 等。模型中包含了场景的语

境信息,即场景中物体的位置信息、尺度信息、物体之间的关系信息以及场景中存在某一物体的概率信息等。与显式的知识表达相比,这类信息多为隐含于模型中的相对信息。场景的语境信息主要分为两类:场景全局信息和空间语境信息。前者表示对场景的总体认识,隐含了该场景中可能存在的物体;后者表示场景中所出现物体之间的关系信息,如草位于天空下方,树木位于可行道路两边等。文献<sup>[62-64]</sup> 利用场景中物体的低层统计特征来建立场景的语境信息模型,而文献<sup>[65]</sup> 则根据场景的高层信息利用条件随机场形成场景中不同物体标记间的语义约束。

### 5.2 视觉算法控制策略

视觉算法控制策略是指移动机器人系统为了使室外自然场景的辨识与理解效果达到最佳而对算法流程进行的选择。当前研究中,计算机视觉算法与机器学习技术的融合为视觉算法控制策略提供了新思路。文献<sup>[14]</sup> 给出了三种视觉控制结构,分别是自顶向下(top-down)、自底向上(bottom-up)和混合结构(hybrid),具体如图 2 所示。

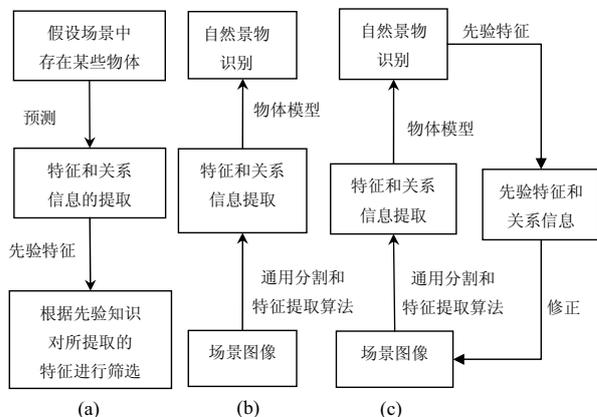


图 2 视觉算法控制策略

(a) 自顶向下 (b) 自底向上 (c) 混合结构

Fig. 2 Vision algorithm strategies

(a)top-down (b)bottom-up (c)hybrid

(1) 自顶向下结构: 基于自顶向下结构的视觉系统也被称为特殊目的视觉系统 (SPVS, Special Purpose Vision System)。该结构对所辨识的场景形成初始假设,即假设场景中某些物体,随后利用学习阶段得到的这些物体的特征信息在该场景中进行验证和筛选。所采用的分割和分类算法根据场景中景物的差异而不同,其辨识的速度快且效果好,但该结构对场景施加了较大的约束,无法对场景中出现的未知物体进行分类辨识。He<sup>[62]</sup>、Shotton<sup>[38]</sup>、Verbeek<sup>[40]</sup> 均采用自顶向下结构。Marti<sup>[66]</sup> 以颜色和纹理作为特征,采用区域生长和区域验证不断循环的方法,并与模糊规则

相结合, 获得了较好的辨识结果。但是由于采用串行区域技术, 时间开销大且无法解决过分割问题。

(2) 自底向上结构: 基于自底向上的视觉系统也被称为通用视觉系统。与自顶向下结构相反, 该方法初始认为场景是未知的。利用分割技术将场景划分为不同的特征区域, 并且所选用的特征维数固定。然后利用已知的物体特征模型与场景的特征区域进行匹配来完成辨识。该结构可以描述场景中的未知区域, 从而增强辨识效果。如 Shotton<sup>[22]</sup> 的 Textonboost 方法即采用自底向上的结构。虽然自底向上结构理论上可以适应具有模糊性和多变性的室外自然场景, 但是该视觉系统结构的构建难度大, 并且很大程度上要依赖于通用特征的选择, 所以往往需要与自顶向下方法相结合。

(3) 混合结构: 混合结构结合了自顶向下结构和自底向上结构的优点, 典型的混合结构有如下两类。第一类是首先基于自底向上结构并采用图像分割算法将场景分为不同区域, 然后在不同区域中利用自顶向下结构对假设存在的物体特征进行搜索验证。第二类是首先基于自顶向下结构对场景中假设存在的物体进行分类辨识, 然后利用自底向上结构对未知区域进行分割描述, 从而实现特征更新或引入新物体特征的目的。其它混合结构大都属于上述两类结构的变形。Bosch<sup>[67]</sup> 首先采用自顶向下的方法生成概率像素图, 提取物体主干并通过区域生长竞争获得已知物体区域。然后再利用自底向上方法分割得到未知区域, 最后根据近邻信息对未知区域进行合并。对于室外复杂自然场景的理解, 一个固定不变的算法策略往往缺少必要的灵活性, 所以采取多策略融合的手段以取得较好的效果。

## 6 自然场景认知

从认知论的角度来考虑<sup>[68]</sup>, 为了能够对自然场景的整体或各局部区域进行有效理解, 首先必须对场景进行有效的描述, 其中特征选择是其重要环节之一。对于室内结构化环境常常用边缘、角点特征加以描述, 而对于室外的自然场景, 由于区域的模糊性往往无法找到类似的规则化特征, 因此特征的选择对自然场景的描述至关重要。文献<sup>[69]</sup> 采用颜色均值、颜色直方图以及纹理等分层变长特征来表征不同地表物质。然而实际环境中颜色易受光照强度的影响, 纹理易受尺度的影响, 所以都不适用于复杂多变的自然环境。受生物视觉系统的启发, 生物视觉皮层上的 V1 细胞具有类似 Gabor 滤波器和高斯拉普拉斯滤波器的作用, 即该类细胞可将视觉信号分解为基频信号分量的线性组合, 由此得到的 texton<sup>[70,71]</sup> 特征在处理自然图像上较为常用。文献<sup>[72,73]</sup> 分别采用局部描述符进行移动机器人室

外自定位和地理信息识别, 所用的局部描述符包括 SIFT<sup>[20]</sup>、PCA-SIFT<sup>[74]</sup>、SURF<sup>[75]</sup> 等, 这些特征对环境的尺度、平移、旋转变换均具有较好的不变性。Lowe<sup>[20]</sup> 受中生物视觉模型启发得到尺度不变描述符 SIFT。该模型基于一种复杂的视觉神经细胞, 认为该细胞对场景中特定取向和空间频率的梯度信息敏感, 这种信息代表着某一场景的不变量, 而 SIFT 正是对该不变量的描述。Serre<sup>[76]</sup> 受生物视觉皮质组织机理<sup>[77]</sup> 的启发给出复杂度渐增的生物学新特征, 该特征可更有效应用于大尺度自然场景图像的辨识和理解。

其次从认知的层次上看, Marr<sup>[6]</sup> 在他开创性的著作中给出了视觉理解过程的三个阶段, 即提取角点、边缘、纹理等基本特征的第一阶段; 获取场景可见部分深度、法线方向和轮廓等信息的第二阶段; 恢复、表示和识别三维物体的第三阶段。早期人们认为对场景的理解过程首先是对场景中各物体的识别, 然后才是对场景整体的认识, 也即自下而上的过程。然而近年来认知领域的发展对这一问题有了新认识, 人们只需少量的信息即可快速对场景做出判断。如当看到树木及灌木, 即可做出森林场景的判断。Li<sup>[78]</sup> 指出人们可以快速理解未知场景而无需对场景中的物体进行仔细辨识。Oliva<sup>[46]</sup>, Greene<sup>[79]</sup> 给出了场景的全局描述特征, 可对多场景进行分类辨识。同时选择性注意机制模拟生物的视觉功能, 为解决大数据量的计算, 提高系统实时性提供了新的思路。

## 7 研究重点和技术发展趋势

### 7.1 研究重点

就当前总体研究现状来看, 所发表的研究成果大多是针对算法的研究, 而能在实际自然场景中加以验证的研究结果则相对有限。因此要使该领域研究工作真正面向实用化, 特别要着重考虑如下研究重点:

1) 多种传感器的协同工作。如何协调视觉、激光测距等不同传感器之间的工作, 以及在不同层次上对多源传感器数据进行融合, 是确保移动机器人在室外自然场景中高效稳定工作的关键。利用激光测距所构建的野外环境模型虽然可提供典型地形环境特征, 但由于对有遮挡的环境(灌木、草丛等)以及地表硬度状况(沙土、砂石、水面、淤泥等)的分析能力有限, 这就需要引入视觉传感器来加强对特定环境的深度判断与理解。视觉传感器很容易区分在高程图上很相似的自然场景, 可避免“漏报”(如只根据测距信息将沼泽地归属到安全区域)和“误报”(如将高草丛归属到危险区域)现象的发生。

2) 系统的集成与优化。为了充分发挥视觉系统的功能,需要有效地对视觉系统的各子模块加以组合,这就是系统集成问题。这些模块具体包括实时图像处理模块、自学习模块、知识库模块以及算法控制策略模块等。显然,即使各子模块的功能很强,但如果没用一个好的系统架构加以组织,系统的整体功能也无法优于各子模块的功能。比如一个易扩展、易更新的全局知识结构要求对内具有高的内聚性,对外具有简单易访问的接口,这样可以方便地为整个系统所利用,从而提高系统的稳定性。系统集成问题可以从软件工程的视角加以考虑,所以利用软件工程学为视觉系统设计良好的架构也是该领域研究的重点。

3) 视觉算法控制策略。自顶向下结构的优点是对假设存在的物体进行辨识的算法效率高,而自底向上结构的优点是可以对场景中出现的未知景物信息进行描述,将二者进行有效结合的混合结构同时具有上述两者的优点。但是如何在实际应用中确定最优的算法执行序列,仍需要进一步加以探索与解决。虽然该问题可从优化角度来认识,但面对室外复杂的场景,一个固定不变的算法流程往往缺少必要的灵活性。

4) 系统实时性能的保证。这需要从系统实用性的角度加以研究。现实中只将场景分割为没有形成语义信息的不同区域,这对移动机器人的规划与导航是毫无用处的。所以将分割和辨识技术相结合,并同时采用不同算法控制策略快速地对场景本身或感兴趣区域进行划分,这才是实际应用中迫切需要解决的问题。虽然国内外很多学者对相关问题进行了长期研究,但如何兼顾场景理解的实时性与辨识效果,从而提高系统的实用性仍是一大难点。而图形处理器单元(GPU)的使用以及分布式计算技术的运用,往往为提高系统的实时性提供辅助手段。

5) 系统对环境中的不确定因素的适应。环境中的不确定因素主要包括:场景中同类物体间辨识的模糊性及不同类物体间辨识的混叠性;移动机器人运动所引起的帧抖动问题;场景整体理解的不确定性等。为了克服上述困难,往往采用场景多尺度表达、多层次特征联合建模、图像变换、帧间冗余信息处理等技术手段。但由于缺少统一的解决框架,使得该问题成为移动机器人自然场景理解研究的技术难点,同时也是该领域未来发展亟待解决的关键问题。

6) 融入视觉注意机制的场景分析辨识<sup>[80]</sup>。视觉注意可分为三种类型:第一类是基于初级视觉,由数据驱动的自底向上的注意,其中最具代表性的是Itti提出的高斯金字塔模型<sup>[81]</sup>。但大多数模型只考虑了显著区域的局部特性,没有考虑整幅图像的整体统计特性<sup>[82]</sup>。第二类是基于高层视觉,与任务、

知识等相关的自顶向下的注意。相比于第一类,此类研究较少。较具有代表性的模型有神经网络模型、统计贝叶斯模型、马尔可夫模型、任务导向模型等。第三类是自底向上和自顶向下处理过程的统计结合<sup>[83]</sup>,但目前感知模型对自顶向下和自底向上的视觉注意之间的关系缺乏明确的计算理解,有待于进一步的研究和建立完整合理的基于注意机制的神经计算理论框架。

## 7.2 技术发展趋势

由于室外自然场景所具有的非结构化与不确定性特点,与传统的视觉系统相比,室外基于视觉的移动机器人系统更要着重考虑运动状态下对环境的自适应性。相关领域的新技术和新思路也不断涌现,以下几点是对该领域技术发展趋势的展望:

1) 视频理解技术。3.2节讨论的图像区域标记方法仅针对图像集之中的单幅静态图像,并且图像之间没有相互关联。而室外移动机器人所要处理的自然场景图像不仅是场景视频中的某一帧,而且各视频帧之间有很高的相关性,因此如何利用帧间相关信息来改善场景辨识效果,从而提高区域标记的实时性是基于单幅图像处理所无法实现的。Brostow<sup>[84]</sup>等人给出了第一个高标准标记的参考视频库CamVid。文献<sup>[85]</sup>使用该视频数据库,从摄像机自运动(ego-motion)过程所获得的三维点云数据中提取场景低维结构特征向量,并将其投影到二维图像平面,进而采用训练过程中得到的随机决策树来辨识二维场景图像中的物体。由于算法没有使用基于外观的描述符(如颜色、纹理等),所以对场景的尺度、天气、光线变化有很强的适应性,同时系统也具有高的实时性。基于场景运动特征的方法从另一个角度来分析处理场景视频,使得该方法成为视频分析与辨识研究领域的一个新热点。

2) 动态知识结构。移动机器人实际运行前要经过费时的训练或者凭借经验来构建规模庞大的知识体系结构,这种知识库所包含的信息相对固定,往往不易于更新。这一缺点使系统无法适应超出训练集以外或各种非经验情况下环境信息的改变。而构建系统动态知识结构的目的在于选择良好的知识表达,如可将显式知识结构和隐式知识模型相结合,进而建立优良的知识框架,使得移动机器人可以与视觉信息相配合,有选择地更新已有知识信息,从而确保系统实际运行的稳定性和可靠性。

3) 对场景信息的自学习。如何通过自学习使机器人能够对环境实现自主适应是非常具有挑战性的研究问题,而将不同的机器学习方法与视觉处理方法相融合正成为该领域的一大发展趋势。现有的与视觉相结合的学习方法都有一定的局限性,如有监

督学习算法无法对场景进行在线学习；半监督学习算法虽然减少了离线的训练量，但是算法往往不稳定；自监督学习也只是针对稳定的感兴趣区域进行辨识。如何通过在线学习机制对自然场景进行深度理解仍然是急需解决的热点和难点问题。

4) 认知学新特征与场景表达的融合。通过借鉴生物视觉模型，可将场景表达与来自认知学新视觉特征进行融合，为场景辨识提供一种新的技术手段。由于传统的图像分割和分类方法较耗时，同时所选用的特征也易受室外随机因素的影响，所以从另一个角度来认识该问题往往受到很大启发。例如提取自然场景中的显著性区域作为自然路标，从而可以避免耗时的图像处理。由此可见认知学新特征与场景表达的融合也是未来该领域的一大发展趋势。

## 8 结束语

基于视觉的室外自然场景的实时辨识与深入理解，是具有实时环境认知与自主环境适应能力的移动机器人系统中最为关键的共性技术之一。要在该领域取得创新性的研究成果，必须兼顾研究的创新性与实用性。多种技术的应用以及不同传感器之间的协作与融合都能有效降低自然场景理解中不确定因素的影响，从而为机器人系统的自主环境适应提供保障，而这一技术的深入发展也必将为机器人自主行为理论的研究注入崭新的活力。

## References

- 1 DeSouza G N, Kak A C. Vision for mobile robot navigation: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(2): 237–267
- 2 Oliva A, Torralba A. Modeling the shape of the scene: a holistic representation of the spatial envelope. *International Journal of Computer Vision*, 2001, **42**(3): 145–175
- 3 Hoiem D. Seeing the World Behind the Image: Spatial Layout for 3D Scene Understanding. Technical Report, CMU-RI-TR-07-28, Robotics Institute, Carnegie Mellon University, USA, 2007
- 4 Heitz G, Gould S, Saxena A, Koller D. Cascaded classification models: combining models for holistic scene understanding. In: Proceedings of Neural Information Processing Systems. Vancouver, Canada: NIPS, 2008.1–8
- 5 Buehler M, Iagnemma K, Singh S. *The 2005 DARPA Grand Challenge: The Great Robot Race*. Berlin: Springer-Verlag, 2007
- 6 Marr D. *Vision*. USA: W. H. Freeman and Company, 1982. 8–38
- 7 Nosofsky R M. Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology*, 1986, **115**(1): 39–61
- 8 Fischler M A, Elschlager R A. The representation and matching of pictorial structures. *IEEE Transactions on Computers*, 1973, **22**(1): 67–92
- 9 Biederman I. Recognition-by-components: a theory of human image understanding. *Psychological Review*, 1987, **94**(2): 115–147
- 10 Pomerleau D A. Knowledge-based training of artificial neural networks for autonomous robot driving. In: Proceedings of Robot Learning. Norwell, MA: Kluwer, 1993. 19–43.
- 11 Jochem T, Pomerleau D, Thorpe C. Vision-based neural network road and intersection detection and traversal. In: Proceedings of IEEE Conference on Intelligent Robots and Systems. Pittsburgh, USA: IROS, 1995. 344–349.
- 12 Bischof W F, Caelli T. Scene understanding by rule evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(11): 1284–1288
- 13 Draper B A, Hanson A R, Riseman E M. Knowledge-directed vision: control, learning, and integration. In: Proceedings of the IEEE. USA: IEEE, 1996, **84**(11): 1625–1637
- 14 Batlle J, Casals A, Freixenet J, Marti J. A review on strategies for recognizing natural objects in colour images of outdoor scenes. *Image and Vision Computing*, 2000, **18**(1): 515–530
- 15 Li F F, Fergus R, Perona P. One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(4): 594–611
- 16 Oliva A, Torralba A. Building the gist of a scene: the role of global image features in recognition. *Progress in Brain Research*, 2006, **155**(1): 23–36
- 17 Oliva A, Torralba A. The role of context in object recognition. *Trends in Cognitive Sciences*, 2007, **11**(12): 520–527
- 18 Larlus D, Jurie F. Combining appearance models and markov random fields for category level object segmentation. In: Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Alaska, USA: CVPR, 2008. 1–7
- 19 Lafferty J, McCallum A, Pereira F. Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: Proceedings of the Eighteenth International Conference on Machine Learning. Williams College, Williamstown, MA, USA: ICML, 2001. 282–289
- 20 Lowe D G. Distinctive image features from scale-invariant key-points[J]. *International Journal of Computer Vision*, 2004, **60**(2): 91–110
- 21 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
- 22 Shotton J, Winn J, Rother C, Criminisi A. TextonBoost for image understanding: multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *International Journal of Computer Vision*, 2009, **81**(1): 2–23
- 23 Fauqueur J, Brostow G J., Cipolla R. Assisted video object labeling by joint tracking of regions and keypoints. In: Proceedings of the Eleventh IEEE International Conference on Computer Vision. Rio de Janeiro, Brazil: ICCV, 2007. 1–7
- 24 Hu Bin, He Ke-Zhong. Applications of computer vision to outdoor mobile robot. *Acta Automatica Sinica*, 2006, **32**(5): 774–784 (胡斌, 何克忠. 计算机视觉在室外移动机器人中的应用. *自动化学报*, 2006, **32**(5): 774–784)
- 25 Wu Dong-Hui, Ye Xiu-Qing. An uncertain knowledge based real time road scene understanding algorithm. *Journal of Image and Graphics*, 2002, **7**(1): 69–74 (吴东晖, 叶秀清. 基于不确定性知识的实时道路场景理解. *中国图象图形学报 (A) 辑*, 2002, **7**(1): 69–74)

- 26 Wang Lu, Lu Xiao-Xia, Cai Zi-Xing. Local salient regions based natural scene recognition. *Journal of Image and Graphics*, 2008, **13**(8): 1594–1600 (王璐, 陆筱霞, 蔡自兴. 基于局部显著区域的自然场景识别. 中国图象图形学 (A) 辑, 2008, **13**(8): 1594–1600)
- 27 Huhns M N, Mendoza B, Ruvinsky A, McCants D. The JIDOKA System for Multiple-sensor Terrain Classification, Technical Report CSE TR-2006-13, Center for Information Technology, University of South Carolina, USA, 2006
- 28 Kelly A, Mandelbaum R, Pilarski T, et al. Toward reliable off road autonomous vehicles operating in challenging environments. *The International Journal of Robotics Research*, 2006, **25**(5): 449–483
- 29 Schafer B H. Security Aspects of Motion Execution in Outdoor Terrain [Ph. D. dissertation], Kaiserslautern Institute of Technology, 2005
- 30 Thrun S, Strohband S, Bradski G., Mahone P, et al. Stanley: The robot that won the DARPA grand challenge. *Journal of Field Robotics*, 2006, **23**(9): 661–692
- 31 Zhang Yu-Jin. *Advances In Video And Image Segmentation*. USA: IRM, 2006. 1–15
- 32 Martin D R, Fowlkes C C, Malik J. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2004, **26**(5): 530–549
- 33 Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis And Machine Intelligence*, 2002, **24**(5): 603–619
- 34 Shi J, Mailik J. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(8): 888–905
- 35 Zhu S, Yuille A. Region competition: unifying snakes, region growing, and bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1996, **18**(9): 884–900
- 36 Yang A, Wright J, Ma Y, Sastry S. Unsupervised segmentation of natural image via lossy data compression. *Computer vision and image understanding*, 2008, **110**(2): 212–225
- 37 Everingham M, Van Gool L, Williams C K I, Winn J, Zisserman A. The PASCAL visual object classes challenge 2007 [Online], available: <http://www.pascal-network.org/challenges/VOC/voc2007/workshop>, 2007
- 38 Shotton J, Johnson M, Cipolla R. Semantic texton forest for image categorization and segmentation. In: Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Alaska, USA: CVPR, 2008. 1–8
- 39 Verbeek J, Triggs B. Region classification with markov field aspect models. In: Proceedings of the 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Minneapolis, USA: CVPR, 2007. 1–8
- 40 Verbeek J, Triggs B. Scene segmentation with conditional random fields learned from partially labeled images. In: Proceedings of Neural Information Processing Systems. Vancouver, Canada: NIPS, 2007. 1–8
- 41 Vailaya A, Figueriredo M, Jain A, Zhang H. Image classification for content-based indexing. *IEEE Transaction on Image Processing*, 2001, **110**(1): 117–130
- 42 Bosch A, et al. A review: which is the best way to organize/classify images by content? *Image and Vision Computing*, 2007, **25**(6): 778–791
- 43 Fan J, Gao Y, Luo H, Xu G. Statistical modeling and conceptualization of natural images. *Pattern Recognition*, 2005, **38**(6): 865–885
- 44 Li F F, Perona P. A bayesian hierarchical model for learning natural scene categories. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA: CVPR, 2005. 524–531
- 45 Bosch A, Zisserman A, Munoz X. Scene classification via pls. In: Proceedings of European Conference on Computer Vision. Graz, Austria: ECCV, 2006. 517–530
- 46 Oliva A, Torralba A. Scene-centered description from spatial envelope properties. In: Proceedings of the Second International Workshop on Biologically Motivated Computer Vision. Tubingen, Germany: BMCV, 2002. 263–272
- 47 Heitz G, Koller D. Learning spatial context: using stuff to find things. In: Proceedings of the European Conference on Computer Vision. Marseille, France: ECCV, 2008. 30–43
- 48 Manduchi R, Castano A, Talukder A, Matthies L. Obstacle detection and terrain classification for autonomous off-road navigation. *Autonomous Robots*, 2005, **18**(1): 81–102
- 49 Huertas A, Matthies L, Rankin A. Stereo-based tree traversability analysis for autonomous off-road navigation. In: Proceedings of the Seventh IEEE Workshops on Application of Computer Vision. Washington, USA: IEEE, 2005. 210–217
- 50 Dima C, Hebert M, Stentz A T. Enabling learning from large datasets: applying active learning to mobile robotics. In: Proceedings of the International Conference on Robotics and Automation. New Orleans, USA: IEEE, 2004. 108–114
- 51 Rosenberg C, Hebert M, Schneiderman H. Semi-supervised self-training of object detection models. In: Proceedings of the Seventh IEEE Workshops on Application of Computer Vision. Washington, USA: IEEE, 2005. 29–36
- 52 Jansen P, Van der Mark W, Van den Heuvel J C, Groen F C A. Colour based off-road environment and terrain type classification. In: Proceedings of the IEEE Conference on Intelligent Transportation Systems. Vienna, Austria: IEEE, 2005. 216–221
- 53 Karlsen R E, Witus G. Terrain understanding for robot navigation. In: Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robotics and Systems. San Diego, CA, USA: IEEE, 2007. 895–900
- 54 Happold M, Ollis M, Johnson N. Enhancing supervised terrain classification with predictive unsupervised learning. In: Proceedings of Robotics: Science and Systems. Philadelphia, USA: RSS, 2006. 1–8
- 55 Sofman B, Lin E, Bagnell J, Vandapel N, Stentz A. Improving robot navigation through self-supervised online learning. In: Proceedings of Robotics: Science and Systems. Philadelphia, USA: RSS, 2006. 1–8
- 56 Stavens D, Thrun S. A self-supervised terrain roughness estimator for off-road autonomous driving. In: Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence. Virginia, USA: UAI, 2006. 13–16
- 57 Dahlkamp H, Kaehler A, Stavens D, Thrun S, Bradski G. Self-supervised monocular road detection in desert terrain. In: Proceedings of Robotics: Science and Systems. Philadelphia, USA: RSS, 2006. 1–7
- 58 Hadsell R, Sermanet P, Ben J, Erkan A, Scoffier M, Kavukcuoglu K. Learning long-range vision for autonomous off-road driving. *Journal of Field Robotics*, 2009, **26**(2): 120–144

- 59 Draper B A, Collins R T, Brolio J, Hanson A R, Riseman E. The schema system. *International Journal of Computer Vision*, 1989, **2**(1): 209–250
- 60 Start T M. Natural Object Recognition. New York: Springer-Verlag, 1992
- 61 Crevier D, Lepage R. Knowledge-based image understanding systems: a survey. *Computer Vision and Image Understanding*, 1997, **67**(2): 161–185
- 62 He X, Zemel R S, Carreira-Perpinan M A. Multiscale conditional random fields for image labeling. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, USA: CVPR, 2004. 695–702
- 63 Murphy K, Torralba A, Freeman W. Using the forest to see the tree: a graphical model relating features, objects and the scenes. In: Proceedings of Neural Information Processing Systems, British Columbia, Canada: NIPS, 2003.1–8
- 64 Torralba A. Contextual priming for object detection. *International Journal of Computer Vision*, 2003, **53**(2): 169-191
- 65 Rabinovich A, Vedaldi A, Galleguillos C. Objects in Context. In: Proceedings of the Eleventh IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil: ICCV, 2007. 1–8
- 66 Marti J, Freixenet J, Batlle J, Casals A. A new approach to outdoor scene description based on learning and top-down segmentation. *Image and Vision Computing*, 2001, **19**(14): 1041–1055
- 67 Bosch A, Munoz X, Freixenet J. Segmentation and description of natural outdoor scenes. *Image and Vision Computing*, 2007, **25**(5): 727–740
- 68 Cai Zi-Xing, Zou Xiao-Bing. Research on environmental cognition theory and methodology for mobile robots. *Robot*, 2004, **26**(1): 87–91 (蔡自兴, 邹小兵. 移动机器人环境认知理论与技术的研究. *机器人*, 2004, **26**(1): 87–91)
- 69 Angelova A, Matthies L, Helmick D, Perona P. Fast terrain classification using variable-length representation for autonomous navigation. In: Proceedings of the 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Minnesota, USA: CVPR, 2007. 567–574
- 70 Julesz B. Textons, the elements of texture perception and their interactions. *Nature*, 1981, **290**(5802): 91–97
- 71 Zhu S C, Guo C, Wang Y, Xu Z. What are textons? *International Journal of Computer Vision*, 2005, **62**(1): 121–143
- 72 Valgren C, Lillenthal A. SIFT, SURF and Seafans: long-term outdoor localization using local features. In: Proceedings of the Third European Conference on Mobile Robots. Freiburg, Germany: ECMR, 2007. 253–258
- 73 Li Y, Lim J H. Outdoor place recognition using compact local descriptors and multiple queries with user verification. In: Proceeding of the 15th International Conference on Multimedia. Augsburg, Germany: ACM Multimedia, 2007. 549–552
- 74 Ke Y, Sukthankar R. PCA-SIFT: a more distinctive representation for local image descriptors. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, USA: CVPR, 2004. 506–513
- 75 Bay H, Ess A, Tuytelaars T, Van Gool L. SURF: speeded up robust features. *Computer Vision and Image Understanding*, 2008, **110**(3): 346–359
- 76 Serre T, Wolf L, Bileschi S, Riesenhuber M, Poggio T. Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **29**(3): 411–426
- 77 Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 1999, **2**(11): 1019-1025.
- 78 Li F F, VanRullen R, Koch C, Perona P. Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, 2002, **99**(14): 9596–9601
- 79 Greene M R, Oliva A. Recognition of natural scenes from global properties: seeing the forest without representing the trees. *Cognitive Psychology*, 2009, **58**(2): 137–179.
- 80 Hurlbert A, Poggio T. Visual information: do computers need attention? *Nature*, 1986, **321**(12): 651–652
- 81 Itti L, Koch C. Computational modeling of visual attention[J]. *Nature Reviews Neuroscience*, 2001, **2**(3) :194–230
- 82 Alvarez G A, Oliva A. The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, 2008, **19**(4): 392–398.
- 83 Long Fu-Hui, Zheng Nan-Ning. A visual computing model based on attention mechanism. *Journal of Image and Graphics*, 1998, **3**(7): 592–595 (龙甫荟, 郑南宁. 一种引入注意机制的视觉计算模型. *中国图象图形学报*, 1998, **3**(7): 592-595)
- 84 Brostow G J, Fauqueur J, Cipolla R. Semantic object classes in video: a high-definition ground truth database. *Pattern Recognition Letters*, 2008, **30**(2): 88–97
- 85 Brostow G J, Shotton J, Fauqueur J, Cipolla R. Segmentation and recognition using structure from motion point clouds. In: Proceedings of the European Conference on Computer Vision. Marseille, France: ECCV, 2008. 44–57



庄 严 大连理工大学自动化系副教授。主要研究方向包括机器人导航, 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 interests include mobile robot navigation, exploration, autonomous environment modeling and cognition. Corresponding author of this paper.)



陈 东 大连理工大学自动化系硕士研究生。2007 年获大连理工大学材料学院材料物理专业学士学位。主要研究方向为机器学习, 室外移动机器人场景感知与理解。E-mail: zaknova@gmail.com

(CHEN Dong Graduate student in the Department of Automation, Dalian University of Technology. He received his bachelor degree from Dalian University of Technology in 2007. His research interest covers machine learning, scene understanding for outdoor mobile robot.)



**王 伟** 大连理工大学信息与控制研究中心主任, 教授. 研究方向包括预测控制, 机器人学及智能控制等.

E-mail: wangwei@dlut.edu.cn

(**WANG Wei** Professor and director of Research Center of Information and

Control, Dalian University of Technology. His research interests include predictive control, robotics, and intelligent control.)



**韩建达** 博士, 中国科学院机器人学重点实验室研究员. 主要研究方向为基于传感器的机器人鲁棒控制系统设计, 多机器人系统, 智能控制系统设计.

E-mail: jdhan@sia.cn



(**HAN Jian-Da** Ph.D. professor at

Robotics Laboratory, Shenyang Institute of Automation, Chinese Academy of Sciences. His research interest covers sensor based robot robust control system design, multi-robot system, and intelligent control system design.)

**王越超** 中国科学院沈阳自动化研究所研究员. 主要研究方向为智能机器人.

E-mail: ycwang@sia.cn

(**WANG Yue-Chao** Professor at Shenyang Institute of Automation, Chinese Academy of Sciences. His main re-

search interest is intelligent robots.)