

基于“原型”的机器人开放式室内场所感知实验研究

朱 博, 戴先中, 李新德, 杨 伟, 陈芳园

(复杂工程系统测量与控制教育部重点实验室(东南大学自动化学院), 江苏 南京 210096)

摘 要: 在基于“原型”的场所感知算法基础上构建了一种场所感知实验平台. 其中, 硬件系统主要以双目立体相机作为环境感知传感器, 以 CPU-GPU (中央处理器-图形处理器) 协同计算单元作为多种任务、复杂逻辑和密集数据的处理单元; 而软件系统基于微软 Robotics Developer Studio 平台构建, 双目视觉服务和场所感知服务为其中两大核心服务, 前者利用 RANSAC (随机采样一致性) 和 GPU 版本 ASIFT (仿射尺度不变特征变换) 等算法在宽视角范围内实现物品识别和位姿估算, 适用于拥挤室内环境, 后者基于双目视觉结果和已有场所感知算法完成场所分类及区域感知并构建 2D 语义地图. 在无人工标注的真实室内环境中开展实验, 实验结果表明, 基于本文实验平台的机器人能够在线鲁棒地感知活动空间中的开放式场所, 实验也在一定程度上验证了已有场所感知算法的可实现性、有效性和实用性.

关键词: 场所分类; 开放式室内场所; 原型理论; 场所感知; GPU; ASIFT; 语义地图

中图分类号: TP242

文献标识码: A

文章编号: 1002-0446(2013)-04-0491-09

Experimental Study on Open Interior-Places Perception of Robot Based on “Prototype”

ZHU Bo, DAI Xianzhong, LI Xinde, YANG Wei, CHEN Fangyuan

(Key Laboratory of Measurement and Control of CSE, School of Automation, Southeast University, Nanjing 210096, China)

Abstract: A place-perception experimental platform is proposed on the basis of the prototype based place perception algorithm. In the hardware system, a binocular rig is used as environment perception sensor, and CPU (central processing unit) and GPU (graphic processing unit) coordinated computing unit is used to process multiple tasks, complex logic, intensive data, and so on. The software system is constructed based on Microsoft Robotics Developer Studio platform. Binocular vision service and place perception service are two core services in it. The former is fit for clustered indoor environment, and realizes object recognition and pose estimation simultaneously in wide viewing angle by using RANSAC (RANDOM Sample Consensus), ASIFT (affine scale-invariant feature transform) -GPU and other algorithms. The latter realizes place categorization and region perception and constructs 2D semantic map based on the results of binocular vision and the existing place perception algorithm. The experiments are conducted in a real interior environment in which no artificial label is used. The experiment results show that the robot can robustly perceive open places in its action space online based on the experimental platform, and realizability, effectiveness and practicability of the existing place perception algorithm are verified to a certain extent.

Keywords: place categorization; open interior place; prototype theory; place perception; GPU (graphic processing unit); ASIFT (affine scale-invariant feature transform); semantic map

1 引言 (Introduction)

人们需要服务机器人更加有意义地推理、规划、实施和完成服务任务, 并且需要其使用人类能够理解的概念与人无缝交互, 因此, 机器人领域研究人员近年广泛地并逐步深入地开展语义层面的研究. 当前, 这方面研究中一个较为活跃的主题是如何使服务机器人对其活动空间进行语义层感知. 众多文

献常使用术语——场所分类 (place classification^[1] 和 place categorization^[2-3] 等) ——提及该主题. 严格意义上, 其内涵不同于场所识别 (place recognition), 场所分类面临的主要挑战之一是如何处理类内存在的差异性^[4]. 从概念层面看, 本文涉及的场所感知问题属于场所分类, 并且包含对某类场所占有空间范围的感知.

基金项目: 国家自然科学基金资助项目 (60804063, 61175091); 江苏省自然科学基金资助项目 (BK2010403); 图像信息处理与智能控制教育部重点实验室开放基金资助项目 (200902); 东南大学优秀青年教师教学、科研资助计划资助项目 (3208001203); 东南大学创新基金资助项目 (3208000501).

通讯作者: 朱博, zhuboseu@163.com 收稿/录用/修回: 2012-09-14/2012-10-31/2013-01-23

“场所 (place)”与“场景 (scene)”概念并不等同, 文 [5] 给出“场所”及“场景”的一种形式化定义, 较为严格地区分了两类. 由文中所给定义可以看出, “场所”实际上由“场景”及“场景”区域间存在的空间关系定义. 场所的空间关系属性在文 [5] 中通过“场景”区域间的相对空间关系体现, 而在本文中, 此属性则通过概念组 (group) 在全局地图上的绝对空间位置给出.

在机器人场所感知这一研究领域中, 传统研究通常以有明确物理边界的房间 (room) 为主要研究对象, 显式地以门或窄通道作为房间划分的线索 [2,4,6] 或者隐式地根据图的谱聚类 (spectral clustering) 生成房间 [7-9], 之后进一步根据其中容纳的物品对房间具体概念进行感知, 这些方法的共同点在于依赖物理分隔物实现场所感知. 然而, 许多人类所涉及的场所概念难以首先通过几何属性感知到 (例如: 教室前部) [10]. 此外, 随着现代开放式室内设计理念的流行, 日常生活中越来越多的功能场所转变为开放或半开放式, 这成为服务机器人进行场所认知所面临的新挑战.

开放式场所感知算法研究近年刚刚起步, 尚未成熟. Vasudevan 等主要提出 4 种方法 [1], 其中, 方法 M1、M2 和 M3 仅考虑场所中的物品, 方法 M4 在 M3 的基础上又考虑了物品间的距离关系, 其工作仅从概念上对场景进行分类, 未考虑以显式形式标示场所概念对应空间区域和体现场所的空间关系属性. 文 [5] 给出的场所感知框架能够实现开放式场所感知, 但遗憾的是, 作者假定事先给定场所特征, 而未进一步详细讨论适用于开放式场所感知的特征形式, 并且作者仅从理论上进行了说明, 没有进行实验验证. 文 [10] 提出一种“上下文依赖的空间区域 (context-dependent spatial region)”, 该概念与本文提到的“开放式场所”非常类似. 作者利用物体实例间的定性空间关系 (拓扑和位置关系) 描述, 采用类比 (analogy) 方法, 并借助于锚点 (anchor point) 实现开放式场所感知, 文 [10-11] 分别以仿真和实验验证所提方法, 仅限于对“教室前部”这类场所进行感知. 文 [12] 提出以抽象场所概念原型来描述场所, 以一种新的相似性度量函数及空间关系评分准则度量测试环境与场所概念原型的相似程度, 实现场所感知全过程. 文 [12] 方法与文 [10] 方法的不同之处在于: 文 [12] 采用物品类别及抽象空间关系原语捕捉场所概念的语义信息, 而不以场所实例作为语义载体进行类推, 其在更为抽象的层面上处理场所类内差异性. 文 [12] 仅以仿真手段验证所提出的开放

式场所感知算法, 而本文从实验角度进一步验证该算法的可实现性、有效性和实用性.

本文构建一种场所感知实验平台并在无人工标注的真实环境中开展实验. 硬件系统主要以双目立体相机作为环境感知传感器, 以 CPU-GPU 协同计算单元处理多任务、复杂逻辑和密集数据; 软件系统则基于微软 Robotics Developer Studio 平台构建, 其中双目视觉服务和场所感知服务为两大核心服务, 前者同时实现物品类别和位姿识别, 后者完成场所类别及区域感知并构建 2D 语义地图.

2 “原型”感知算法简介 (Overview of the “prototype” perception algorithm)

我们认为, 在开放式室内环境中, 人们首先根据记忆系统中的概念原型把具有功能、目的、属性和关系相关性的一些环境物聚类感知为一个具有概念的特定组; 然后, 进一步感知该组对应的空间区域. 机器人可以模拟人的这种行为进行场所感知.

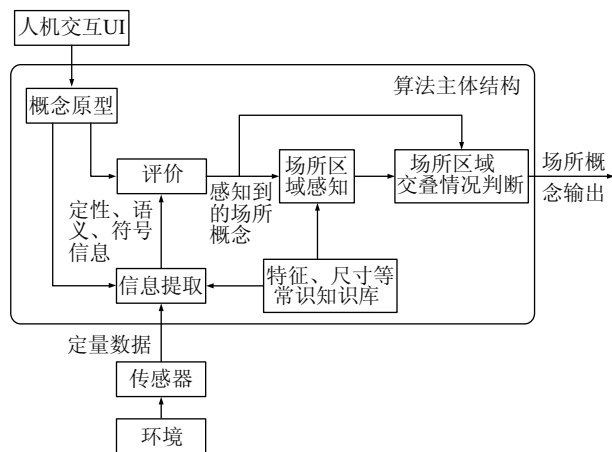


图1 基于原型的场所感知算法整体框图

Fig.1 The framework of place perception algorithm based on prototype

算法整体框图如图1所示. 执行过程为: 首先, 参数化开放式场所概念的原型模型 (通过人机交互或其它手段), 该模型形式化表达如下:

$$\langle N_0, F_0, \Omega_0, S_0 \rangle$$

其中, N_0 为场所概念的名称, F_0 为场所概念的特征物品集合, Ω_0 为与集合 F_0 元素一一对应的权重系数集合, S_0 为描述物品间空间关系的谓词集合. 其次, 探索环境的过程中, 机器人通过底层传感器数据获取所需大粒度定性信息及高层抽象信息, 进而评价当前环境与已有概念原型的相似度, 如果达到一定程度, 则初步感知出对应的抽象场所概念. 最后, 根据知识库知识在内部地图中标示感知到的场所的最

确定区域, 初步完成场所区域感知; 若多个场所区域间无交叠或者少交叠, 则生成并输出感知到的场所概念, 否则取消区域标识且不生成场所概念. 这样, 即完整实现了对场所的概念感知和区域感知过程.

有关概念原型模型、原型匹配所用相似性度量函数及空间关系评分准则等的详细内容参见文 [12], 不再赘述. 下文实验平台围绕感知算法对相关信息的需求构建.

3 实验平台设计 (Design of the experimental platform)

基于“原型”的场所感知算法建立在较为高层的定性、语义和符号信息上, 从而使得算法本身与底层传感器类型、数据形式和实现形式解耦, 原型知识构建和扩充过程亦不需要底层数据参与, 场所感知过程甚至不要求底层传感器提供精确信息 (例如: 只需粗略估计物体位姿). 基于该算法的上述特点, 对任意机器人平台, 倘若能够从底层传感器的现有数据中提取所需信息, 便可为之增加开放式场所感知功能. 文 [13] 由此提出基于物联网技术实现开放式场所感知的构想. 可见, 利用前述感知算法能够构建的开放式场所感知实验平台可以不唯一.

尽管实验手段存在非唯一性, 但本文以常见双目立体相机作为环境信息采集传感器, 以带有里程计及方位传感器的机器人作为载体构建的实验平台, 具有一定的通用性与代表性. 其中, 双目视觉传感器及相应软件服务提供场所感知算法所需的环境物品类别和位姿信息, 场所感知软件服务进一步根据物品信息以感知算法分析机器人当前所处环境蕴

含的场所概念.

3.1 硬件系统及实验装置设计

系统硬件构成见图 2.

本文所用机器人本体具有带编码器的差动驱动单元、方位传感器、蓝牙通信模块、机器人控制器、双目摄像机和两组独立无线视频发射机, 而其本身不具备复杂算法处理能力, 相关算法由机器人服务器处理完成. 机器人服务器通过蓝牙模块和两组无线视频接收机与机器人本体交换信息. 模拟视频信号经由两路 PCI 视频采集卡转换为数字信号, 进一步由 CPU-GPU 协同计算单元处理. CPU-GPU 协同计算架构保证复杂信息处理过程的实时性. 其中, CPU 主要负责多种服务任务调度、GPU 任务调度、低运算量数值运算 (场所感知运算、物品位姿估计等)、复杂逻辑运算和用户界面显示等, GPU 用以并行实现基于图像 ASIFT^[14] 特征的特征提取和匹配运算. 场所感知过程得到的物品和场所概念以音频方式告知用户.

机器人本体的实物图见图 3. 本文使用 Mindstorms NXT 2.0 系列套装, 它由乐高公司同麻省理工学院共同研发^[15], 能够提供小型、可定制机器人搭建平台. 机器人控制器 (整合蓝牙通信模块) 和差动驱动单元分别采用 9797 教育套装中的 NXT 模块和伺服电机模块. 9797 套装的部分配件搭配 9695 配件库构建机器人底盘及其它结构框架. 方位传感器采用 HiTechnic NMC1034 模块. 双目摄像机及无线视频发射机采用航模领域常见的 FPV (first person view, 第一人称视角) 摄像机及视频发射装置. 为使机器人从合适高度观察室内环境, 采用纯铝支架支撑双目摄像机, 使之离地高度约 0.56 m.

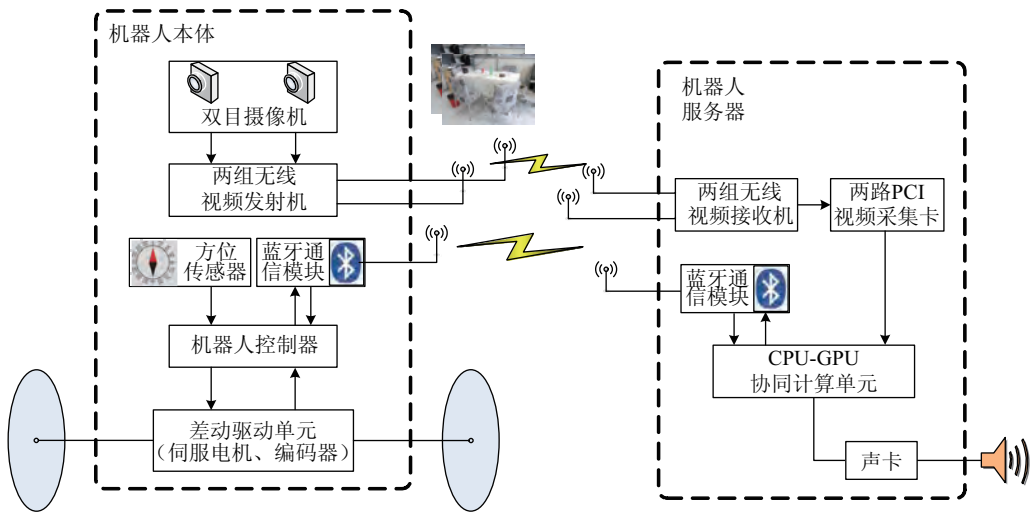


图 2 硬件系统框图

Fig.2 The framework of the hardware system

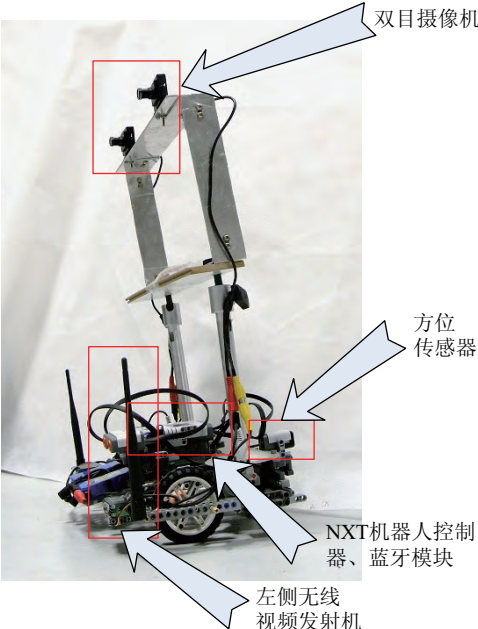


图 3 机器人实物图

Fig.3 The photo of the real robot

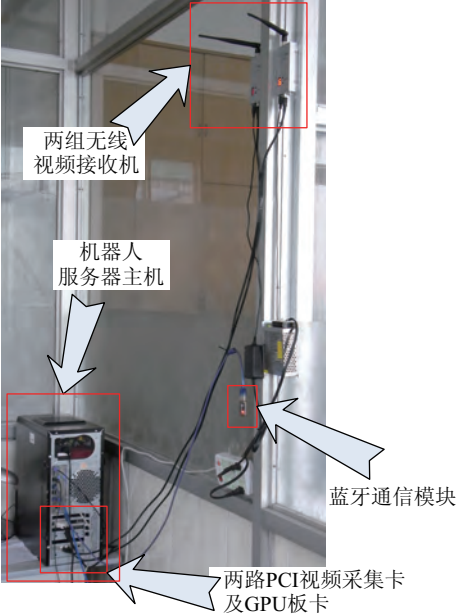


图 4 机器人服务器实物图

Fig.4 The photo of the robot server

机器人服务器实物图见图 4。两组无线视频接收机与发射端一一对应，使用不同频道。蓝牙模块带有 USB2.0 接口，蓝牙版本为 2.0。PCI 视频采集卡为微视 V221 双路采集卡，直接获取 RGB24 格式、640×480 分辨率数字图像。服务器主机中 CPU 为 Intel i5-2320(4 核，3.0 GHz)，内存容量为 4GB (DDR3-1333)，GPU 为 NVIDIA GTX560Ti (1.64 GHz/2.004 GHz，2GB，GDDR5 显存)。

3.2 软件系统设计

本文的软件系统是基于微软 Robotics Developer

Studio (MSRDS) [16] 机器人开发平台构建的。DSS (decentralized software service, 分散式软件服务) 是 MSRDS 中重要编程模型，提供了轻量级、面向状态的服务模型。DSS 机制有助于以模块化形式实现各种功能，它所实现的机器人应用程序具有鲁棒、高性能、松耦合等优点。通常采用多个 DSS 实现多个功能模块，再通过软件组态形成应用系统。系统中各功能模块间能够实时传递信息、相互协作完成不同任务以及实现无阻塞异步数据处理等。

基于 DSS 服务架构的软件系统结构图见图 5。

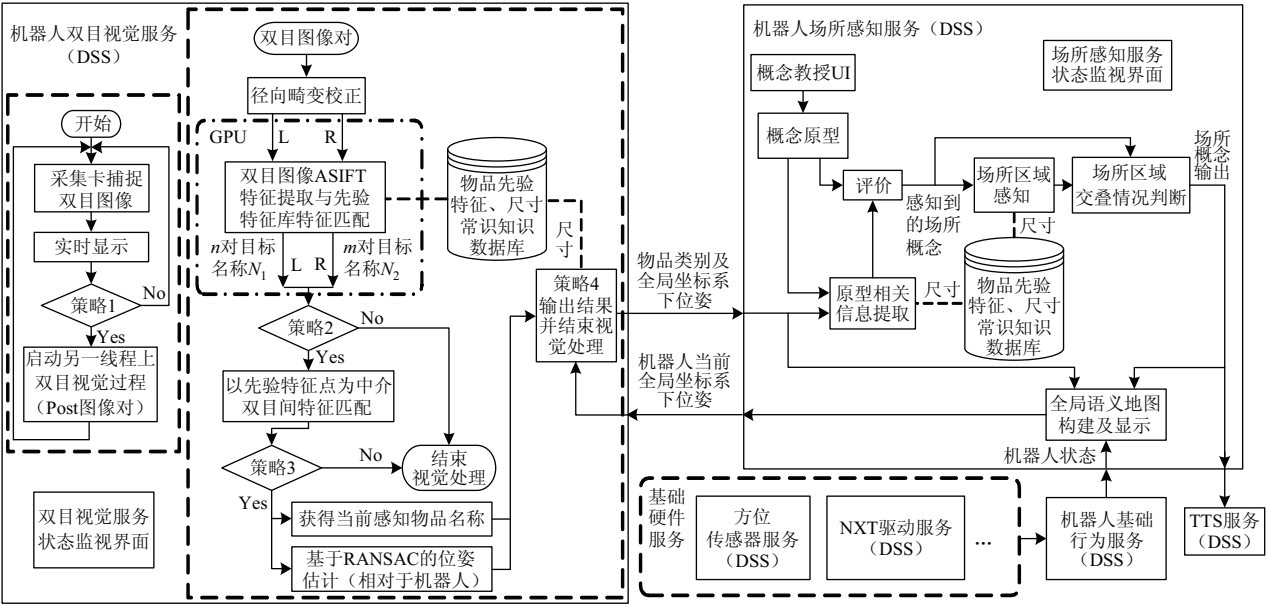


图 5 软件系统结构图

Fig.5 The structure diagram of the software system

软件系统结构图中, 多个 DSS 服务松耦合地构成整个机器人场所感知软件系统. 方位传感器服务、NXT 驱动服务等对硬件层作了适当抽象, 便于上层服务使用. 机器人基础行为服务向多个硬件服务传递指令、查询状态、接收通知, 进而提供机器人应用层功能, 包括手动/自主控制机器人行为、监视机器人运行状态等. 本文使用了 Bodurov^[17] 提供的基础行为服务包, 在其基础上增加了“依规划行动序列运行”行为模式和机器人运行轨迹记录及显示功能.

机器人双目视觉服务和场所感知服务为本文两大核心服务, 前者实现对物品的感知(包括物品类别和位姿), 后者在前者基础上实现完整的场所感知(包括场所类别及对应区域), 同时构建环境的 2D 语义地图. 两者均使用物品常识知识库, 由其提供相关物品必要的先验特征及尺寸知识(来自国家和行业标准^[12]). 物品先验特征为其先验正面图像上的 ASIFT 特征, 该特征能够保证在较少先验知识下, 机器人从大视角范围内对物品进行识别并估计其位姿. 机器人每发现新物品或新场所则通过 TTS (text to speech, 从文本到语音) 服务以语音形式通知用户.

双目视觉服务包括两个主要部分: 双目图像捕捉过程和双目视觉处理过程(图 5 双目视觉服务中虚线框所示). 双目图像捕捉过程中, 每隔一定时间捕捉并实时显示图像对, 当满足策略 1 条件(于下一节介绍)时, 图像对被发送到双目视觉处理线程处理. 图像对发送过程利用 MSRDS 中另一核心机制——CCR (concurrency and coordination runtime, 并发协作运行时), 该机制具有易于管理异步操作、处理程序并发和使用并行硬件等优势. 由于来自 FPV 摄像机的图像具有较大畸变, 图像对被发送到另一线程后, 首先对其进行径向畸变矫正, 以便后续处理过程能够得到相对准确的结果. 然后, 分别对双目图像提取 ASIFT 特征, 并与物品先验特征数据库进行匹配, 输出最匹配物品类别名称及匹配点对数目. 由于该过程数据密集且计算复杂, 使用 CPU 计算相当耗时, 不能满足实时性要求, 因此采用 GPU 实现. 当双目中匹配结果满足策略 2 条件(于下一节介绍)时, 在双目间进行特征点匹配: 如果分别来自双目中的特征点以同一先验特征点为匹配点, 则这两个特征点匹配, 这种双目间匹配方式能够保证匹配点对尽可能落于目标物上. 当满足策略 3 条件(于下一节介绍)时, 能够以较高概率确定目标物品的类别名称, 并进一步进行位姿估计.

当不满足策略 2 或者策略 3 条件时, 结束视觉过程. 基于 RANSAC^[18] 的物品位姿估计过程可得到目标物品在机器人坐标系下的位姿, 该过程将在下一节介绍. 利用策略 4 (于下一节介绍) 综合视觉处理结果并将结果输出给场所感知服务.

场所感知服务中, 主要实现了基于“原型”的开放式场所感知算法^[12]. 需要指出的是, 其中的场所概念教授界面并不唯一, 可采用多种方式从用户获取某场所概念描述, 例如: 采用对话框形式, 询问相关描述; 也可以采用文 [19] 提出的手绘界面, 捕捉用户给出的场所草图中的原型信息; 将来甚至可以采用基于自然语言理解的方式捕捉语言中的场所概念信息. 另外, 为监视服务的运行状态, 分别为两个核心服务设计了状态监视界面.

4 相关策略及算法 (Related strategies and algorithms)

4.1 双目视觉服务相关策略

双目视觉服务使用多种策略保证系统可靠、有效运行. 当软件系统被移植到其它机器人平台时, 可调整相应策略以适应新工况.

策略 1 当且仅当机器人停止移动 t 秒 ($t > 0$) 且双目视觉处理程序空闲时, 启动新的双目视觉处理过程.

策略 2 当且仅当 $n > n_{\text{threshold}}$, $m > m_{\text{threshold}}$ 且 $N_1 = N_2$ 时进行双目间匹配. 其中, 各符号含义见图 5, $(\cdot)_{\text{threshold}}$ 为阈值.

策略 3 当且仅当双目间匹配点对数量大于阈值(阈值通常大于等于 4)时, 进行后续位姿估计等过程.

策略 4 根据机器人在全局坐标系下的当前位姿将机器人坐标系下物品位姿转换到全局语义地图上, 当且仅当全局语义地图上以新识别物品位置为圆心、物品较大先验形状外接圆半径为半径的圆区域范围内没有同类物品时, 保留新识别物品.

4.2 径向畸变矫正算法

对于广角 FPV 摄像机镜头, 必须进行径向畸变矫正.

设摄像机坐标系下的任意 3D 空间点为 $\mathbf{X}_c = [X_c \ Y_c \ Z_c]^T$, 规范化为 $\mathbf{x}_n = [x \ y]^T = [X_c/Z_c \ Y_c/Z_c]^T$, 其到畸变中心半径为 $r(x, y)$, 径向畸变模型为 $f(r)$. 另设已知畸变图像上任意点像素坐标为 $\mathbf{X}_p$ (齐次坐标), 径向畸变矫正目标为: 求 $\mathbf{X}_p$ 对应的非畸变图像上坐标 $\mathbf{X}_{p, \text{und}}$. 通常采用迭代法求解径向畸变矫正, 而本文通过摄像机立体标定容易得到摄像机内

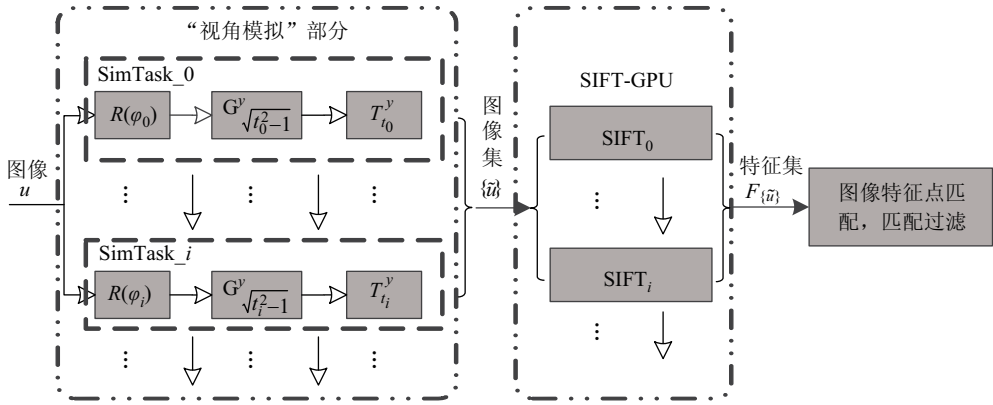


图 6 ASIFT-GPU 结构图
Fig.6 The structure diagram of ASIFT-GPU

参阵 $\mathbf{K}$ ，因此无需迭代求解便可快速实现径向畸变矫正。根据式 (1) 可以得到 $\mathbf{X}_{p,ud} \rightarrow \mathbf{X}_p$ 的逆映射：

$$\mathbf{X}_p = \mathbf{K} [f(r) \ f(r) \ 1]^T \cdot \mathbf{K}^{-1} \mathbf{X}_{p,ud} \quad (1)$$

其中， r 由后两项 $\mathbf{K}^{-1} \mathbf{X}_{p,ud}$ 结果得到。获得逆映射后利用双线性插值即得到径向畸变矫正结果。

4.3 ASIFT-GPU 并行算法

ASIFT 特征提取算法是 Morel 和 Yu 于 2009 年提出的一种完全仿射不变特征提取方法 [14]。ASIFT 算法为机器人目标识别带来如下好处：

(1) 只需要物品的一幅先验图像（通常为正面图像），而无需多角度先验图像，使得先验知识获取过程在一定程度上得到简化，特别地，对大体积家具来说此优势更为明显；

(2) 使得机器人能够从相对于目标物品更宽视角范围内发现并识别目标，对于较为拥挤的室内环境，使得识别过程更加便利。

尽管 ASIFT 算法具有完全仿射不变性能，但是低运算效率使得其难以直接应用。本文设计了 ASIFT 算法的 GPU 版本（简称 ASIFT-GPU），使之能够在低成本 PC 上准实时实现。

ASIFT-GPU 结构图见图 6，图中只标示出算法主要环节。

整个 ASIFT-GPU 由 3 大部分构成：视角模拟、SIFT-GPU 和特征匹配。其中，阴影框所示运算环节均由 GPU 并行实现；空心箭头表示两相邻运算由 CPU 先后启动但非严格串行，其中核函数有可能并发执行（本文 GPU 具有 2.1 计算能力）；实心箭头表示相邻环节严格串行，图中表达式含义详见文 [14]。

以图 6 结构实现 ASIFT 算法并适当优化，算法效率得以大幅提升，CPU 版本同 GPU 版本耗时比较见表 1。其中，图像均规范化到 400×300 分辨率，使用标准 Adam 图对 [14] 测试。

表 1 ASIFT 算法 CPU 与 GPU 版本比较
Tab.1 The comparison of ASIFT-CPU and ASIFT-GPU

	特征提取时间（视角模拟+SIFT）/ms	特征匹配时间/ms
CPU 版本 （关闭 OpenMP）	2273+4544=6817	3202
CPU 版本 （开启 OpenMP）	830+1275=2105	1960
GPU 版本	77.347+128.445=205.792	65.02
加速比	33/10	49/30

可见 ASIFT-GPU 能够达到准实时在线处理要求。

4.4 基于 RANSAC 的位姿估计算法

原型描述通常涉及物品间空间关系，这些关系需要基于物品位姿判定。室内大量家具具有“面”结构，而利用双目获取的目标表面深度信息可以重建出一个几何平面，这可作为物理家具“面”结构的拟合模型。通过将该平面的法向量投影到 2D 地图平面，并对平面模型内点（inlier）求平均，即可获得目标物的 2D 位姿估计。

基于双目间匹配特征点，利用三角计算容易获得对应空间点的集合 $\{P_s\}$ 。由于物理家具的“面”结构并非理想平面并且存在传感器噪声和其它不确定因素，因此 $\{P_s\}$ 中大部分点分布于某理想平面周边，少数点远离该平面。本文采用 RANSAC 鲁棒估计方法求此 $\{P_s\}$ 上拟合平面。算法流程图见图 7。

5 实验结果（Experimental results）

实验条件：

1) 感知算法参数设置 [12]： $\alpha = 1$ ；“round($\cdot, \cdot$)”、“half-round($\cdot, \cdot$)”和“in-front-of($\cdot, \cdot$)”关系的接受范围分别为 [1.5, 3]、[0.5, 2.5] 和 [0.75, 1]；区域交叠判断阈值 $p_1 = p_2 = 0.5$ 。

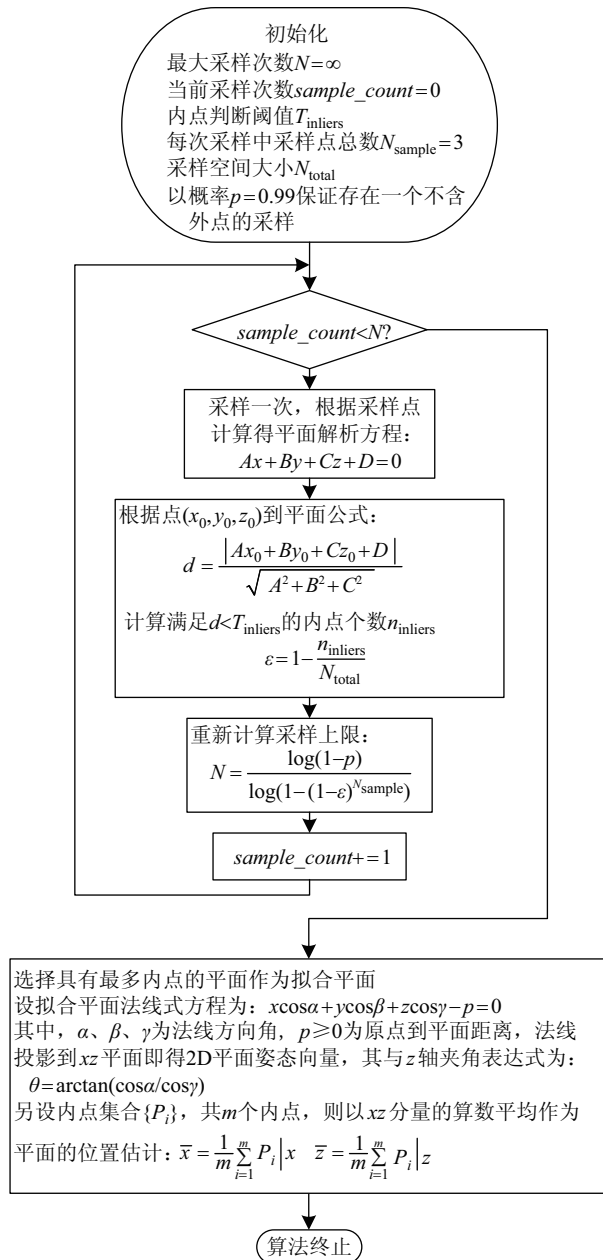


图 7 基于 RANSAC 的位姿估计算法流程图

Fig.7 The flow diagram of pose estimation algorithm based on RANSAC

2) 由于机器人活动范围较小, 底层未采用复杂定位算法, 仅依赖里程计和方位传感器信息确定机器人当前姿态。

3) 机器人可由手动控制或者根据已规划动作序列在环境中移动, 同时在线进行场所感知。需要指出的是, 由于本文视觉算法未考虑同一视野内多目标识别及位姿估计, 因此, 机器人观察环境时需保证单一目标物尽量占据大部分视野。

4) 先验知识库中保存各物品正面图像特征和标准尺寸上下限。

5) 通过人机交互界面获得场所概念原型如下:

$\langle diningroom, \{f_0(chair) = 4, f_0(table) = 1\},$
 $\{\omega_0(chair) = 0.8, \omega_0(table) = 0.8\},$
 $\{round(chair, table)\}\rangle$
 $\langle livingroom, \{f_0(tv) = 1, f_0(sofa) = 1\},$
 $\{\omega_0(tv) = 0.8, \omega_0(sofa) = 0.8\},$
 $\{face-to-face(tv, sofa)\}\rangle$

上述原型的含义是: “餐厅” 通常由 4 把椅子围绕 1 张桌子构成, 两种物品对生成“餐厅”概念的影响均为 0.8; “客厅” 通常由沙发和与其相向摆放的电视构成, 两种物品对生成“客厅”概念的影响均为 0.8。

5.1 物品识别及位姿估计

已知原型中, 桌椅间定性关系和物品位置有关, 而电视与沙发间定性关系与物品位置和姿态均有关, 因此本节以电视位姿估计为例展示位姿估计结果。为验证算法鲁棒性, 将电视与电视机柜组合作为识别目标, 其不具有客观物理平面, 只具有大致“面”结构。

表 2 给出基于 ASIFT 特征的双目在线匹配及位姿估计结果。其中, 分别从差别较大的 3 个视角观察目标, 目标均被识别且得到位姿估计值, 这说明选用 ASIFT 特征有助于从宽视角范围内识别目标。位姿估计结果均投影到机器人局部坐标系 xz 平面, 位置以 (x, z) 表示, 直观起见, 姿态以平面法向量投影与 z 轴负方向夹角度数表示, 物体朝向视野右侧为负, 反之为正。表中基准值用于衡量位姿估计结果, 对于 3D 物品, 由于难以测量其在某坐标系下真实位姿, 简单起见, 本文以电视机柜中心到机器人局部坐标系原点距离作为位置评价基准, 以机器人底盘纵轴与电视机柜前沿夹角的余角作为姿态向量评价基准。以表 2 中 (a) 情况为例具体说明: 特征匹配前双目图像已经过径向畸变矫正; 左目、右目检测到的 ASIFT 特征点数分别为 5552、5209; 左目、右目与先验特征库匹配获得的特征点对数目分别为 185、214; 双目间匹配得到的特征点对数目为 63; 表中相对误差小于 10%, 这在工程中通常可以接受, 事实上, 由于场所感知算法基于定性空间关系设计, 因此相对误差接受范围可进一步放宽。

由实验结果可知, 双目视觉服务能够在宽视角范围内同时实现物品类别和位姿识别, 可应用于较为拥挤的室内环境。

5.2 符合原型描述的环境

由多种真实家具构成本文实验环境, 其中包括餐桌、餐椅、电视机柜 (其上电视由显示器包装纸箱

表 2 双目匹配及位姿估计结果
Tab.2 The results of binocular matching and pose estimation

视角情况 (双目特征点数分别为: LF、RF; 双目与先验特征库匹配结果: LFMced、RFMced 双目间匹配结果: LRMced)	位姿估计结果	基准值	相对误差
(a) 正视目标 LF = 5552、RF = 5209 LFMced = 185、RFMced = 214、LRMced = 63	(133 mm, 1281 mm), 0.099°	1330 mm, 0°	3.17%, -
右侧观察目标 LF = 4784、RF = 4377 LFMced = 80、RFMced = 63、LRMced = 18	(287 mm, 1280 mm), 47.289°	1380 mm, 52°	4.94%, 9.06%
左侧观察目标 LF = 4909、RF = 4944 LFMced = 54、RFMced = 61、LRMced = 14	(59 mm, 1423 mm), -53.676°	1370 mm, -55°	3.96%, 2.41%

代替)和沙发等,家具上不使用任何人工标识物,各种物品由机器人在线识别. 需要指出的是,本文实验所用物品均具有相对丰富的材质表面,市面上同类家具存在另一种情况——家具表面为纯色或者缺乏纹理特征,采用基于 ASIFT 特征的方法对这类物品识别会失效,此时应当采取其它特征(如轮廓特征或直线特征等)对物品进行识别,所选取特征要尽量保证机器人能够在大视角范围下对物品识别并确定位姿. 采用其它特征进行物品识别的讨论超出本文范畴,不再赘述,本文基于 ASIFT 特征的识别方法能够处理一类真实家具的识别问题.

如图 8 所示真实室内环境,物品摆放符合开放式餐厅和客厅布局,机器人根据上述原型对开放式客厅与餐厅进行感知.

图 9 以全局语义地图形式给出感知结果,其中圆圈代表机器人,黑点构成的不规则曲线为机器人的行走路径,物品所占空间范围以国家标准中最小尺寸形成的方框标识,机器人识别出的物品名称和位置分别以字符串及空心圆点标出,物品姿态由代表物品的方框姿态反映出. 由语义地图可知,机器人能够在线感知出开放式“餐厅”和“客厅”(实验过程中机器人以语音方式告知用户),两者所占最确定区域由闭合多边形在语义地图上显式标出. 场所感知算法的相关数据如下:当前环境与餐厅和客厅原型中特征物品集合的相似度分别为 $s_{\alpha}(\cdot, \cdot) = 0.98$ 和 $s_{\alpha}(\cdot, \cdot) = 1$, 其空间关系评分分别为 $\mathcal{ANL}_{x\text{-round}}(\cdot, \cdot) = 2.25$ 和 $\mathcal{ANL}_{\text{face-to-face}}(\cdot, \cdot) = 1$. 值得注意的是,尽管餐厅原型中有 4 把椅子而机器人在真实环境中只观察到 3 把,但观察到的环境与原型知识在一定程度上保持一致,不影响场所概念感知

结果,场所感知算法的更多鲁棒性验证参见文 [12]. 综上,基于本文提出的场所感知实验平台,机器人能够在真实环境中完成对共存开放式场所的感知.



图 8 开放式餐厅和客厅
Fig.8 The open dining room and living room

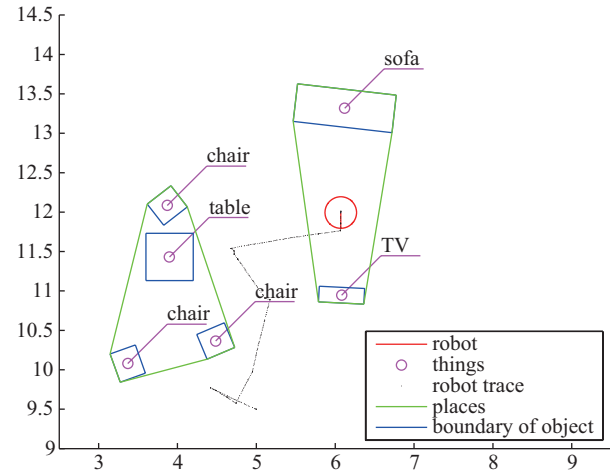


图 9 场所语义地图
Fig.9 The semantic map of places

5.3 不符合原型描述的环境

如图 10 所示另一真实室内环境, 其中物品摆放较为杂乱。所用原型知识不变, 机器人对环境进行感知。



图 10 杂乱室内环境

Fig.10 The disordered indoor environment

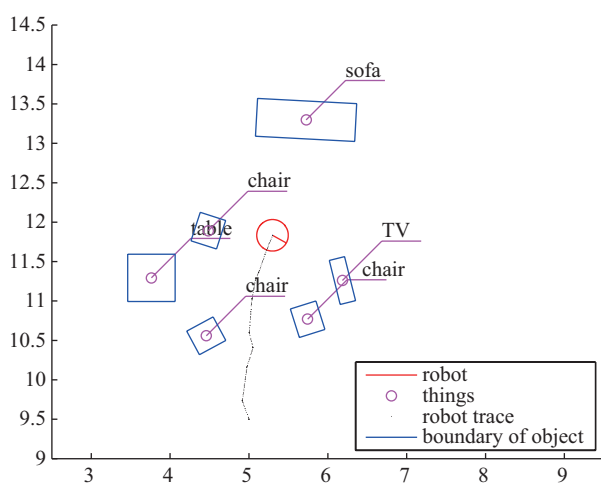


图 11 全局语义地图

Fig.11 The global semantic map

同样以全局语义地图给出感知结果 (见图 11)。此时, 场所感知算法的相关数据如下: 当前环境与餐厅和客厅原型中特征物品集合的相似度分别为 $s_{\alpha}(\cdot, \cdot) = 0.98$ 和 $s_{\alpha}(\cdot, \cdot) = 1$, 其空间关系评分分别为 $\mathcal{ANL}_{x\text{-round}}(\cdot, \cdot) = 1$ 和 $\mathcal{ANL}_{\text{face-to-face}}(\cdot, \cdot) = 0$ 。可见, 虽然机器人观察到环境中的物品, 但是由于物品间空间关系不符合原型描述, 因此机器人未感知出原型知识对应的场所。

6 结论 (Conclusion)

本文在前期开放式室内场所感知算法理论研究基础上, 进一步对其进行实验研究, 提出一种场所感知实验平台, 在无人工标注的真实环境中开展实验。实验结果表明机器人能够在线、鲁棒地利用双目识别环境物品类别、估计物品位姿, 并感知活动空间中存在的开放式场所, 同时构建场所语义地图。实验在一定程度上验证了已有算法的可实现性、有效性和实用性。

本文实验系统在理论和应用方面还需要进一步研究完善, 例如: 需要研究更丰富的空间关系原

语及相应感知算法, 研究多场所区域交叠时分割重组算法, 整合物品分类器以便实现复杂原型中多物品在线识别, 研究多目标双目识别及同时位姿估计算法, 研究复杂外形物品的实时在线位姿估计算法等。

实验过程得到 ASIFT 研究小组中 Yu 和 Limare 两位学者以及英伟达工程师 Wang 的无私帮助, 在此对他们表示由衷感谢!

参考文献 (References)

- [1] Vasudevan S, Siegwart R. Bayesian space conceptualization and place classification for semantic maps in mobile robotics[J]. Robotics and Autonomous Systems, 2008, 56(6): 522-537.
- [2] Pronobis A, Jensfelt P. Hierarchical multi-modal place categorization[C/OL]//5th European Conference on Mobile Robots. [2012-08-25]. <http://cogx.eu/data/cogx/publications/PronobisECMR2011.pdf>
- [3] Swadzba A, Wachsmuth S. Indoor scene classification using combined 3D and gist features[C]//The 10th Asian Conference on Computer Vision (ACCV 2010). Berlin, Germany: Springer-Verlag, 2011: 201-215.
- [4] Pronobis A. Semantic mapping with mobile robots[D]. Stockholm, Sweden: KTH Royal Institute of Technology, 2011.
- [5] Pronobis A, Sjo K, Aydemir A, et al. A framework for robust cognitive spatial mapping[C]//14th International Conference on Advanced Robotics. Piscataway, NJ, USA: IEEE, 2009: 1-8.
- [6] Pronobis A, Martinez Mozos O, Caputo B, et al. Multi-modal semantic place classification[J]. International Journal of Robotics Research, 2010, 29(2/3): 298-320.
- [7] Zivkovic Z, Booij O, Krose B. From images to rooms[J]. Robotics and Autonomous Systems, 2007, 55(5): 411-418.
- [8] Valgren C, Lilienthal A. Incremental spectral clustering and seasons: Appearance-based localization in outdoor environments [C]//IEEE International Conference on Robotics and Automation. Piscataway, NJ, USA: IEEE, 2008: 1856-1861.
- [9] 吴皓, 田国会, 陈西博, 等. 基于机器人服务任务导向的室内未知环境地图构建 [J]. 机器人, 2010, 32(2): 196-203.
Wu H, Tian G H, Chen X B, et al. Map building of indoor unknown environment based on robot service mission direction[J]. Robot, 2010, 32(2): 196-203.
- [10] Klenk M, Hawes N, Lockwood K. Representing and reasoning about spatial regions defined by context[C/OL]//2011 AAAI Fall Symposium. El Segundo, CA, USA: AI Access Foundation, 2011: 154-161. [2012-08-25]. <https://www.aaai.org/ocs/index.php/FSS/FSS11/paper/download/4143/4543>.
- [11] Hawes N, Klenk M, Lockwood K, et al. Towards a cognitive system that can recognize spatial regions based on context[C/OL]//26th AAAI Conference on Artificial Intelligence. El Segundo, CA, USA: AI Access Foundation, 2012: 200-206. [2012-08-25]. <http://www.aaai.org/ocs/index.php/AAAI/AAAI12/paper/download/5061/5139>.
- [12] 朱博, 戴先中, 李新德. 基于“原型”的机器人开放式室内场所感知算法 [J]. 模式识别与人工智能, 2012, 25(1): 1-10.
Zhu B, Dai X Z, Li X D. Open interior-places perception algorithm of robot based on prototype[J]. Pattern Recognition and Artificial Intelligence. 2012, 25(1): 1-10. (下转第 512 页)

- [96] Konolige K, Bowman J. Towards lifelong visual maps[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2009: 1156-1163.
- [97] Walcott-Bryant A, Kaess M, Johannsson H, et al. Dynamic pose graph SLAM: Long-term mapping in low dynamic environments[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2012: 1871-1878.
- [98] Koppula H S, Anand A, Joachims T, et al. Semantic labeling of 3D point clouds for indoor scenes[C]//25th Annual Conference on Neural Information Processing Systems. New York, USA: Curran Associates Inc., 2011.
- [99] Bao S Y, Savarese S. Semantic structure from motion[C]//IEEE

Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos, USA: IEEE Computer Society, 2011: 2025-2032.

作者简介:

梁明杰 (1986-), 男, 博士生. 研究领域: 环境建模, 自主导航, 概率推理.

闵华清 (1956-), 男, 教授, 博士生导师. 研究领域: 智能机器人, 智能软件, 自动系统.

罗荣华 (1975-), 男, 副教授, 硕士生导师. 研究领域: 机器视觉, 场景理解, 概率模型.

(上接第 499 页)

- [13] 朱博, 戴先中, 李新德. 一种基于物联网技术的场所感知系统初探[J]. 计算机科学, 2012, 39(2): 216-221.
- Zhu B, Dai X Z, Li X D. Preliminary study on a kind of place perception system based on IOT technology[J]. Computer Science, 2012, 39(2): 216-221.
- [14] Morel J-M, Yu G S. ASIFT: A new framework for fully affine invariant image comparison[J]. SIAM Journal on Imaging Sciences, 2009, 2(2): 438-469.
- [15] MIT Media Lab. Lifelong Kindergarten[OL]. [2012-08-25]. <http://llk.media.mit.edu/projects.php?id=135>.
- [16] Microsoft Robotics. Microsoft Robotics Developer Studio[OL]. [2012-08-25]. <http://www.microsoft.com/robotics/#About>.
- [17] Bodurov V. Using Microsoft Robotics Developer Studio with Lego Mindstorms NXT Robot[OL]. (2010-08-10) [2012-08-25]. <http://www.simple-talk.com/dotnet/.net-framework/using-microsoft-robotics-developer-studio-with-lego-mindstorms-nxt-robot/>.
- [18] Fischler M A, Bolles R C. Random sample consensus: A paradigm for model fitting with applications to image analysis

and automated cartography[J]. Communications of the ACM, 1981, 24(6): 381-395.

- [19] Zhu B, Dai X Z, Li X D, et al. Place concept teaching through sketch map for robot place perception based on prototype mechanism[C]//6th International Conference on Intelligent Systems and Knowledge Engineering: vol.124. Berlin, Germany: Springer-Verlag, 2011: 79-89.

作者简介:

朱 博 (1981-), 男, 博士生. 研究领域: 空间定性推理, 机器人交互与导航, 语义地图创建.

戴先中 (1954-), 男, 博士, 教授. 研究领域: 复杂控制理论, 机器人控制, 电力系统控制, 测量与信号处理.

李新德 (1975-), 男, 博士, 副教授. 研究领域: 智能机器人, 机器感知, 信息融合, 不确定推理和机器视觉.