

基于物体间支撑语义关系的室内场景彩色深度图像分割

陈昊天¹, 郑 阳², 张钰桐², 孙凤池^{2†}, 黄亚楼²

(1. 南开大学 计算机学院, 天津 300071; 2. 南开大学 软件学院, 天津 300071)

摘要: 场景理解是智能自主机器人领域的一个重要研究方向, 而图像分割是场景理解的基础. 但是, 不完备的训练数据集, 以及真实环境中的罕见情形, 会导致在图像分割时存在先验知识不完备的情况, 进而影响图像分割的效果. 因此, 提出在彩色深度(RGB-D)图像上使用抽象的支撑语义关系来解决多样的物体形态所面对的先验知识不完备问题. 在先验知识不完备情况下, 针对自底向上的图像分割过程中被过度分割出的物体块, 首先对物体块间的支撑语义关系进行建模并计算其支撑概率, 然后构造能够度量场景总体稳定性的能量函数, 最后通过Swendsen-Wang割(SWC)随机图分割算法最小化该能量函数的值, 将物体块间的支撑概率转化为强支撑语义关系并完成物体块合并, 实现先验知识不完备情况下的图像分割. 实验结果证明, 结合支撑语义关系的图像分割能够在先验知识不完备的情况下, 将同一物体被过度分割的部分重新合并起来, 从而提升了图像分割的准确性.

关键词: 自主机器人; 场景理解; 图像分割; 先验知识; 支撑语义关系

引用格式: 陈昊天, 郑阳, 张钰桐, 等. 基于物体间支撑语义关系的室内场景彩色深度图像分割. 控制理论与应用, 2019, 36(4): 579 – 588

中图分类号: TP242 文献标识码: A

DOI: 10.7641/CTA.2018.70825

Indoor red green blue-depth segmentation based on object-object supportive semantic relationships

CHEN Hao-tian¹, ZHENG Yang², ZHANG Yu-tong², SUN Feng-chi^{2†}, HUANG Ya-lou²

(1. College of Computer Science, Nankai University, Tianjin 300071, China;

2. College of Software, Nankai University, Tianjin 300071, China)

Abstract: One of the most important research fields for intelligent autonomous robots is scene understanding which requires a foundation of image segmentation. However, it usually faces a condition of incomplete prior knowledge due to incomplete training dataset and uncommon situations in the real world, which affects the segmentation quality. Therefore, a method that solves the incomplete prior knowledge problem caused by the diversity of objects is presented, which is based on extracting abstract supportive semantic relationships in red green blue-depth (RGB-D) images. This method aims at the over-segmented object parts during the bottom-up image segmentation process with incomplete prior knowledge. First, the supportive semantic relationships between object parts are modelled by calculating supportive probabilities. Second, an energy function is built to measure the total stability of the entire scene. Last, Swendsen-Wang cuts (SWC) algorithm is implemented to minimize the energy function so that the supportive probabilities between object parts are converted into strong supportive semantic relationships. According to the relationships, the object parts are merged to accomplish image segmentation with incomplete prior knowledge. The experiments show that the over-segmented parts are merged into one single object with incomplete prior knowledge by simply introducing supportive semantic relationships to image segmentation. With our method more precise image segmentation results are achieved.

Key words: autonomous robots; scene understanding; image segmentation; prior knowledge; supportive semantic relationship

Citation: CHEN Haotian, ZHENG Yang, ZHANG Yutong, et al. Indoor red green blue-depth segmentation based on object-object supportive semantic relationships. *Control Theory & Applications*, 2019, 36(4): 579 – 588

收稿日期: 2017-11-13; 录用日期: 2018-05-22.

†通信作者. E-mail: fengchisun@nankai.edu.cn; Tel.: +86 13821501302.

本文责任编辑: 苏剑波.

国家自然科学基金项目(61573196), 天津市自然科学基金项目(16JCYBJC18300, 15JCYBJC18800), 中国民航信息技术科研基地开放课题基金项目(CAAC-ITRB-201606)资助.

Supported by the National Natural Science Foundation of China (61573196), the Natural Science Foundation of Tianjin (16JCYBJC18300, 15JCYBJC18800) and the Open Project Foundation of Information Technology Research Base of Civil Aviation University of China (CAAC-ITRB-201606).

1 引言

近年来,智能机器人技术的发展突飞猛进,然而机器人与人之间仍存在着巨大的交互障碍,使其难以全面进入普通人的日常生活.室内服务机器人能够根据传感器进行避障和导航,但是对场景所蕴含的意义缺乏认知,难以在理解场景意义的基础上主动执行用户的意图.上述问题的解决受制于场景理解研究的发展程度,在场景理解技术支持下,服务机器人将获得与人类相容的场景语义标签、空间位置和内涵属性等信息,从而促使其更加自主地完成室内服务任务,并且可以与人类进行自然交互.因此,场景理解成为了智能自主机器人领域的一个重要研究方向.

场景理解具有3层结构^[1],分别对应传感器信息处理层(即处理低级视觉、激光等信息)、语义分析层(即处理语义特征)和认知层(即处理高级抽象语义标签).其中,中层的存在是为了减小低层和高层之间的语义鸿沟^[2].场景理解必须建立低层传感器信息与高层场景语义标签之间的映射关系,而图像分割是获得该映射关系的第一步,它是将图像分离成不同部分的过程,使得每部分清晰地包含一个物体,从而更加易于分析和后续处理^[3],而分割的质量将直接影响场景理解的结果.

在移动机器人场景认知领域,由于无监督的图像分割不包含语义信息^[4],难以保证其结果符合人的认知,因此多采用有监督分割方法,而有监督分割的先验知识有着不完备的缺点:例如,事先获取的数据集无法包含机器人所有的服务场景;标注的尺度和细节因不同情景而异;特征选择不完善,同时又受到了数据和标签的制约.基于单个物体图像分割方法对人提供的先验知识的精确程度和广泛程度非常敏感,即如果在数据标注过程中出现错标、漏标、或在实际应用时环境发生变化,那么机器人就不能正常完成对图像的分割,从而限制了已有方法的适应能力和推广能力.对于自底向上的图像分割,在先验知识缺失、不完备或有误的情况下,很难正确地完成图像分割任务,分割结果中存在被过度分割出的物体块,它们是同一物体的不同部分.

室内环境中的各种物体并不是孤立存在的,物体与物体之间往往存在着丰富的语义关系.这些语义关系能够为服务机器人在图像分割与场景理解过程中提供丰富的信息.文献[5]使用条件随机场(condition-al random field, CRF)综合考虑了局部外观、几何特征以及场景元素之间的相互关系来实现物体分割.文献[6]以平面为元,通过建立语法规则的方式来解码空间结构,得到物体、表面、矩形和空间关系的层次化树状图.文献[7]通过统计的方式学习数据集中物体的共

现关系和位置关系来自动随机生成室内虚拟场景.文献[8]使用潜在语义对分层形变部件模型(hierarchical deformable part model, HPM)进行改进,得到场景层次化的语义分解.文献[9]通过提取空间特征和空间关系特征来训练空间模型,在根据投票获得最终的空间关系指标,再将该指标转化为定性的空间关系,最终基于概率方法完成自顶向下的空间推理.上述方法虽然考虑了场景内物体之间的语义关系,但并没有针对先验知识不完备的情况.

支撑语义关系是一类特殊的语义关系:弱支撑语义关系存在于不同物体之间,它描述了物体整体保持稳定所需的物体之间的支撑,是可分割的;强支撑语义关系存在于同一物体内部被过度分割的物体块之间,它描述了物体块保持稳定所需的物体块之间的支撑,是不可分割的.因此,本文提出基于物体间支撑语义关系的彩色深度(red green blue-depth, RGB-D)图像分割方法,通过识别物体间抽象的支撑语义关系,来弥补丰富的物体形态所造成的先验知识不完备,并利用强支撑语义关系将同一物体被过度分割的部分重新合并起来,从而提高RGB-D图像中物体分割的准确性.流程图如图1所示,首先将RGB-D图像初步分割后所得到的物体块的点云进行下采样并转化为几何体,然后估计每个几何体内的属性,通过该属性计算物体块间的支撑概率,完成对支撑语义的建模,之后设计单个物体的不稳定性函数和全局的能量函数,并利用随机图分割的方法最小化该能量函数,通过参数调节支撑语义关系的强弱,使得算法仅识别物体块间的强支撑语义关系,最后根据该关系将物体块合并为同一物体,以完成更准确的图像分割.

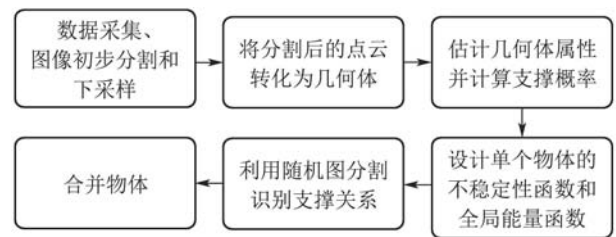


图1 基于物体间支撑语义关系的RGB-D图像分割流程图

Fig. 1 Flow chart for RGB-D segmentation based on supportive semantic relationships

本文的创新点在于对支撑语义关系进行建模并根据其强弱区分物体内部物体块之间的支撑语义关系和不同物体间的支撑语义关系,但图像分割的最终结果不仅仅基于几何和物理关系,还结合了初分割中的颜色、纹理等特征,引入支撑语义关系是为了弥补先验知识不完备的情况.下面与相关工作进行简要比较.文献[10]对场景点云数据进行几何和物理推理,利用Swendsen-Wang割(Swendsen-Wang cuts, SWC)算法

通过补充三维体元和估计体元总体稳定性的方式来进行物体分割. 与本文方法相比, 该方法并没有利用颜色信息, 并且没有针对先验知识不完备的情况. 文献[11]将点云进行几何体拟合, 并在几何体的基础上进行几何推断, 然后通过构造基于物理关系的能量函数并进行最小化求解来实现物理关系推断, 完成分割过程. 与本文方法相比, 该方法没有对支撑语义关系进行统一的建模. 文献[12]对物体的位姿稳定性进行建模, 通过模拟人所产生的场和自然灾害所产生的场来检测潜在可能坠落的物体. 文献[13]和文献[14]采用三维体元的方向梯度直方图 (histogram of oriented gradient, HOG) 特征, 通过训练自适应增强 (adaptive boosting, Adaboost) 分类器来获得物体危险指数, 进而评估环境中的风险. 与之相比, 本文中的支撑语义关系与风险评估的意义和侧重点不同. 文献[15]通过研究主动轮廓模型在图像分割中的多阶段决策过程, 构建了蚁群求解算法, 利用最优路径的搜索问题求解图像分割, 以获取精确的图像轮廓. 但该方法不适用于RGB-D图像.

2 支撑语义关系的建模

对物体块间支撑语义关系进行建模. 首先进行RGB-D数据采集, 并根据已有先验知识进行图像初步分割, 形成物体块, 然后通过下采样减少每个物体块内所需处理点云的数量, 之后进行几何体拟合与属性估计, 进一步提取物体块特征的统计量, 最后根据该统计信息计算物体块之间的支撑概率, 该概率度量了物体块之间存在支撑语义关系的可能性.

2.1 数据采集和图像初步分割

利用搭载Kinect的TurtleBot 2机器人同时采集彩色图像信息和深度图像信息(如图2所示), 然后使用文献[16]中的方法对RGB-D图像进行预处理和初步分割: 计算表面法向量; 根据场景主表面对图像进行旋转使得地面和墙面的数据点垂直于 xy 平面; 使用Watershed算法进行无监督的超分割; 利用训练好的边缘分类器, 将超分割图像分块自底向上进行初步合并. 以Kinect的成像中心为坐标原点, 光轴为 z 轴, 重力的反方向为 y 轴建立右手坐标系, 坐标轴单位为米, 并将每一帧深度图像转化为各自坐标系下的点云数据, 由于图像分割以帧为单位, 不需要进行不同坐标系下的变换. 文献[16]方法为有监督的分割方法, 相对于实验采集到的数据, 其训练数据中的先验知识是不完备的.

2.2 下采样

由于Kinect所采集的点云数据量大, 直接处理全部数据点计算速度慢, 而且考虑到Kinect传感器在感知物体边缘时会产生数据点不准确的情况, 因此使用下采样的方法来减少点云数据规模和降低边缘数据

点对后续步骤的影响. 经过对比, 实验中对点云中的数据以每5行、每5列为间隔进行一次采样, 在保留原始数据信息的情况下尽可能地减少计算量和边缘误差.



图2 搭载Kinect的TurtleBot 2机器人
Fig. 2 TurtleBot 2 loaded with Kinect

2.3 几何体拟合

点云数据只包含表面轮廓, 不能为支撑语义关系的建模提供体积等三维信息; 将点云进行抽象的描述能够更好地获取物体块属性, 从而计算出支撑语义关系所必要的定量信息; 此外物体块会因遮挡而产生不规则的点云数据, 需要将其补充. 因此将分割和下采样后每个物体块的数据点进行几何体拟合. 由于室内环境的结构规整, 且大部分物体为人造物体, 所以物体本身或其组成成分(物体块)皆可采用立方体对物体进行几何体拟合. 拟合并非局限于待分割物体块必须为长方体, 当物体块为圆柱体或其他形状时, 仍可采用立方体拟合来估计其支撑属性.

1) 提取立方体的主平面方向.

由于点云数据刻画了物体的表面, 因此进行几何体拟合时, 立方体中的一对表面应当与点云数据中的最大平面相平行, 称该方向为主平面方向. 故对物体块点云中全部的数据点进行随机抽样一致性算法 (random sample consensus, RANSAC)^[17]平面拟合, 找到点云中最大平面, 并将其法向量作为拟合立方体的主平面方向.

2) 提取立方体的次平面方向.

同理, 立方体所剩的4个平面中, 另外一对表面应当与点云数据中去掉与最大平面相拟合的点后所剩下的点的次大平面相平行, 称该方向为次平面方向. 故去掉最大平面邻域范围内的点, 对剩下的点进行RANSAC平面拟合, 约束该平面法向量垂直于主平面方向, 找到点云中次大平面, 并将其法向量作为拟合立方体的次平面方向.

若去掉第1个平面邻域范围内的点后的点云中的点不足以计算次大平面,则说明物体块扁平,此时考虑拟合的立方体应当尽量贴近物体块的点云表面.故通过RANSAC寻找原点云中垂直于最大平面的最窄的方向,约束该平面法向量垂直于主平面方向,并将其法向量作为拟合立方体的次平面方向.

3) 进行立方体拟合.

立方体的最后一对平面的法向量应当垂直于主平面方向和次平面方向.故通过计算上述步骤中的两个平面的法向量的外积,来得到立方体的最后一对平面的法向量.

由于拟合立方体应当尽可能地代表物体块点云的物理属性,因此应在能够包含大部分数据点的情况下,使其体积最小.故在已知3个法向量的前提下,通过RANSAC方法寻找包含60%数据点且体积最小的立方体.

由于RANSAC方法具有随机性,因此将上述几何体拟合重复10次,并以体积最小的立方体作为对物体块点云的几何体拟合.

2.4 几何体属性抽取

为了对物体块之间的支撑语义关系进行建模并度量其支撑概率,对每个物体块所拟合的几何体统计如下属性:

1) 体积.

由于点云数据只包含表面轮廓,因此物体块的体积很难直接用数据点进行估计,故采用拟合后的立方体体积来估计,定义为

$$\text{vol} = \text{volume of the box}, \quad (1)$$

该定义在一定程度上补全了点云数据中空的部分.

2) 质量.

由于点云数据只包含表面轮廓,并且其疏密程度与距离传感器远近有关,因此物体块的质量很难直接用数据点进行估计,故使用拟合后的立方体体积和密度来估计,定义为

$$m = \rho \cdot \text{vol}, \quad (2)$$

这里 ρ 为密度参数,假设物体密度均匀.

3) 镜面对称度.

镜面对称度将用来衡量物体块的平衡程度,进而影响该物体块与其他物体块是否存在支撑语义关系.定义为

$$ms^j = \sum_p \text{dist}(p, ap^j), \quad (3)$$

其中: $ap^j(j=1,2,3)$ 表示3个对称面; $\text{dist}(\cdot)$ 是距离函数,距离在对称面法向量方向为正,反之为负; p 为物体块所对应点云数据中的点.即,物体的3个镜面

对称度分别是点云中的点到各对称面的距离和,而对称面由立方体对称面得到.

4) 底面倾斜度.

底面倾斜度将用来衡量物体块翻倒的可能性,进而影响该物体块与其他物体块是否存在支撑语义关系.定义为

$$ba = \text{angle}(bf, z), \quad (4)$$

其中: $\text{angle}(\cdot)$ 为求角度函数, bf 为物体块所拟合几何体的底面, z 为 z 轴法向量;即,底面倾斜度由底面与 z 轴的夹角得出,该夹角越大,物体块的倾斜程度就越严重.

5) 翻倒力矩.

翻倒力矩将用来衡量物体块翻倒的倾向程度,进而影响该物体块与其他物体块是否存在支撑语义关系.定义为

$$t = m \cdot g \cdot \sin(ba) \cdot \text{dist}(c, bf), \quad (5)$$

其中 c 为物体块所拟合几何体的质心点,即,翻倒力矩为重力在质心处对底面的力矩.

2.5 支撑语义关系度量与支撑概率

采用支撑概率来建模物体块之间存在支撑语义关系,该概率衡量了物体块间之间支撑的强弱,即相互支撑的可能性.支撑概率越大,则物体块间的支撑语义关系越强,相关的物体块属于同一物体的不同部分的可能性就越高.支撑概率为下一章定性识别物体块间的强支撑语义关系提供了定量的数学基础.

一物体对另一物体存在支撑语义关系的概率定义为

$$\begin{cases} q_e = q_e^{\text{area}} \cdot \max_v(q_e^{ms_v} \cdot q_e^{ba_v} \cdot q_e^{t_v}), \\ q_e^{\text{area}} = 1 - \exp\left\{-\frac{1}{T_i} \cdot \frac{\text{area}(if)}{\text{area}(isf)}\right\}, \\ q_e^{ms_v} = 1 - \prod_{j=1}^3 \exp\left\{-\frac{|ms_v^j|}{T_s}\right\}, \\ q_e^{ba_v} = 1 - \exp\left\{-\frac{|ba_v|}{T_b}\right\}, \\ q_e^{t_v} = 1 - \exp\left\{-\frac{|t_v|}{T_t}\right\}, \end{cases} \quad (6)$$

其中: $\text{area}(\cdot)$ 为求面积函数, if 为物体立方体被其他物体立方体表面所截的相交截面, isf 为该截面所在的物体立方体表面, ms_v^j , ba_v 和 t_v 分别为物体块 v 的3个镜面对称度、底面倾斜度和翻倒力矩, T_i , T_s , T_b 和 T_t 为权值.可以看出,支撑概率由4部分概率组成, if 与 isf 的面积比反映了潜在支撑物对该物体产生支撑作用的可能性,面积比越大,支撑语义关系越强,而不稳定性随着 ms_v^j , ba_v 和 t_v 的增大而增大,因此支撑概率也越大.

3 基于支撑语义关系的图像分割

上一章已对物体块间支撑语义关系进行了建模, 用支撑概率来度量存在支撑语义关系的可能性. 本章将要设计能量函数, 并使用随机图分割的方法来识别强支撑语义关系, 存在强支撑语义关系的物体块应当属于同一物体, 从而将物体块合并以完成更准确的图像分割.

3.1 物体块集合的不稳定性函数

将物体块集合视为单个物体, 构造其势能函数和不稳定性函数.

考虑物体的重力势能与其高度和质量成正比, 构造单个物体的势能函数

$$G(O_i) = h_{O_i} \cdot \sum_{v \in O_i} m_v, \quad (7)$$

其中: O_i 为物体块集合, m_v 为物体块 v 的质量, h_{O_i} 为物体块集合 O_i 最低点的高度. 显然, 物体块的总质量越大或者高度越高, 其势能越大.

考虑物体的不稳定性除了势能外, 还包括翻倒所释放的能量, 构造单个物体的不稳定性函数

$$\begin{cases} E(O_i) = \mu_g \cdot \text{ratio}_{O_i} \cdot G(O_i) + m_{O_i} \cdot \Delta h_{O_i}, \\ \Delta h_{O_i} = h_{O_i}^{\text{current}} - h_{O_i}^{\text{resting}}, \end{cases} \quad (8)$$

其中: m_{O_i} , $h_{O_i}^{\text{current}}$ 和 $h_{O_i}^{\text{resting}}$ 为物体块集合 O_i 拟合的立方体的总质量、当前高度以及按底面放平时的高度, μ_g 为权重; 而比率为

$$\text{ratio}_{O_i} = \min_v \frac{\text{area}(bf_{v_l})}{\text{area}(if_{v_l, v})}, \quad (9)$$

其中: $v_l \in O_i$ 为 O_i 中高度最小的物体块, bf_{v_l} 和 $if_{v_l, v}$ 分别为它的底面和与它存在潜在相交关系的其他物体块 v 的相交截面. 即 ratio_{O_i} 是物体块集合 O_i 最下方的物体块 v 的底面积与其支撑面积的比, 它代表了整个物体集合释放重力势能的可能性. 当最下方的物体块的支撑面较小时, 物体块集合的不稳定性也越大.

3.2 全局能量函数

构造单帧 RGB-D 图像上所有物体块集合的全局能量函数. 全局能量函数度量了某种物体分割结果下场景的总能量, 不但应包含每个物体块集合的物体的不稳定性, 还应拒绝过度合并物体块, 其定义为

$$E(L|G) = \sum_{O_i \in L} (E(O_i) + F(O_i)), \quad (10)$$

其中: $E(L|G)$ 表示在图 G 下以标签 L 进行物体分割时的全局能量函数, l 为某一标签, O_i 为属于同一标签的一个物体. $F(O_i)$ 为惩罚项, 定义为

$$F(O_i) = \mu_1 f_1(O_i) + \mu_2 f_2(O_i), \quad (11)$$

其中: f_1 为属于同一标签的子物体个数, 即由于已经

进行了初步分割, 属于同一标签的子物体数量不宜太多; f_2 为上方物体与下方物体的体积比, 即下方物体体积远小于上方物体体积时, 不宜赋予同一标签; μ_1 和 μ_2 为权重.

全局能量函数的值越小, 所对应分割中的每个物体块集合的稳定程度就越高.

3.3 基于 SWC 算法的图像分割

SWC 算法^[18-20]是属于马尔可夫链蒙特卡罗(Markov chain monte carlo, MCMC)的一种算法, 它通过在随机图上构造一条马尔科夫链并定义其转移函数来完成对其稳定分布的计算. 将 SWC 算法应用于图像分割, 在这里, 以物体块的几何体为节点 v 、支撑语义关系为边 e 、支撑概率 q_e 为边概率构造随机图, μ_e 表示随机图中边的状态, on 为开启, off 为关闭, 而随机图的连通性对应了场景中物体块的分割状态. 使用 SWC 算法在随机图上反复迭代寻找全局能量函数的最小值, 此时所对应的随机图的边的开启状态, 就是物体块之间的强支撑语义关系.

SWC 算法的流程为:

- 1) 循环, 直到全局能量函数稳定, 当前状态为 $\pi = (V_1, V_2, \dots, V_n)$.
- 2) 选择节点 $v \in V_l$, 令 $V_0 \leftarrow \{v\}$, $C \leftarrow \emptyset$.
- 3) 循环, 直到 $C \cap C(V_0, V_l \setminus V_0) = C(V_0, V_l \setminus V_0)$,
 - ① 任取 $e = \langle s, t \rangle \in C(V_0, V_l \setminus V_0)$, $s \in V_0$, $t \in V_l \setminus V_0$;
 - ② 以概率 q_e 随机开启边 e , 即令 $\mu_e \leftarrow \text{on}$, 否则关闭边 e , 即令 $\mu_e \leftarrow \text{off}$;
 - ③ 若 $\mu_e = \text{on}$, 则令 $V_0 \leftarrow V_0 \cup \{t\}$, 否则, 令 $C \leftarrow C \cup \{e\}$.
- 4) 以概率 $q(c_{V_0} = l' | V_0, \pi)$ 对 V_0 赋予随机新标签 l' .
- 5) 以概率 $\alpha(\pi \rightarrow \pi')$ 随机接受新标签和新分割.

算法中, π 为当前随机图的分割状态, 共分割成 n 个集合 V_1, V_2, \dots, V_n . 不断迭代, 随机选择一个节点 v , 设 v 属于第 l 个集合 V_l , 而后不断以概率 q_e 对节点 v 进行连通生长, 形成新的子集 V_0 , 其与原集合 V_l 所差的边称为割边(cuts). 以概率 $q(c_{V_0} = l' | V_0, \pi)$ 赋予 V_0 新的随机标签 l' , 计算其转移概率 $\alpha(\pi \rightarrow \pi')$, 并以该概率接受新的分割状态 π' . 迭代的终止条件是全局能量函数的最小值在一定迭代次数中保持稳定. 其中, 新标签概率 $q(c_{V_0} = l' | V_0, \pi)$ 定义为均匀分布. 转移概率定义为

$$\alpha(\pi \rightarrow \pi') = \min\left(1, \frac{q(\pi' \rightarrow \pi)}{q(\pi \rightarrow \pi')} \cdot \frac{E(L'|G)}{E(L|G)}\right), \quad (12)$$

其中 $q(\pi' \rightarrow \pi)$ 和 $q(\pi \rightarrow \pi')$ 由所有割边的概率 q_e 计算得出。 $E(L'|G)$ 和 $E(L|G)$ 分别为不同分割下的全局能量函数。

最终,当算法迭代趋于稳定时,分割状态 π 的边状态就是物体块间的强支撑语义关系.通过调节边概率 q_e 和转移概率 $\alpha(\pi \rightarrow \pi')$ 中的参数,控制支撑语义关系的强弱,即降低强支撑语义关系的能量,提高对弱支撑语义关系的惩罚.将存在强支撑语义关系的物体块作为同一物体的不同部分加以合并,实现更精确的RGB-D图像分割.

4 实验结果

使用Kinect在实验室的室内环境进行RGB-D图像采集,共有18个场景,包含墙壁、门、地面、灭火器、箱子、柜子、书桌、水壶等11个室内常见物体.实验平台为MATLAB,操作系统为Ubuntu 16.04 LTS, CPU为i7-6700HQ,程序采用单线程.

初步分割算法利用边缘分类器,将超分割图像分块自底向上进行初步合并^[16],由于训练数据来自于公开数据集^[16],与本文所采集的数据特性存在差异,并且受光照、阴影等因素的影响,形成了先验知识不完备的情况,因此分类器的分割结果仍然存在大量被过度分割的物体.先验知识不完备与选择的方法无关,是必然的结果.

以相邻的两个超分割图像分块是否应该合并作为正例和负例,表1给出了先验知识不完备情况下初步分割方法的精确率、召回率和准确率.可以看出,由于先验知识不包含实验室场景内的物体,初步分割将墙壁、柜子等物体分割为多个部分,因此其精确率和召回率不高.

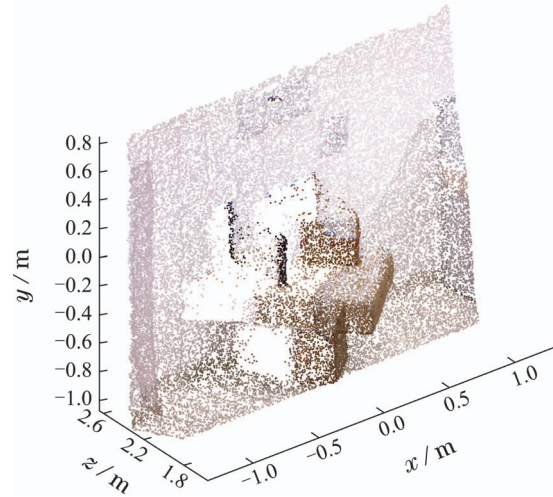
表1 本文方法与初步分割对比结果

Table 1 Comparison between initial segmentation and our method

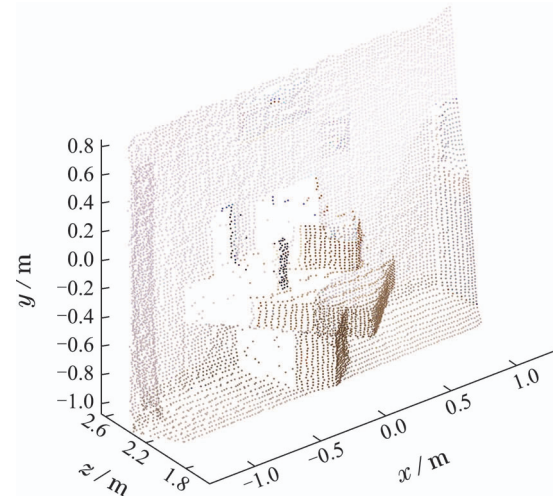
方法	精确率/%	召回率/%	准确率/%
初步分割	77.33	85.10	99.84
本文方法	77.55	86.32	99.84

4.1 下采样与几何体拟合实验结果

下采样的目的是为了减少点云数据规模和降低边缘数据点对后续步骤的影响.图3是一组5行5列下采样前后的点云图像对比,可以看出,在采样前(如图3(a)所示),点云图像密集,但物体边缘有大量离群点,它们增加了几何体拟合的误差;在采样后(如图3(b)所示),点云图像稀疏,但保留了物体的轮廓,且物体边缘的离群点减少,由于几何体拟合是以大部分数据点为拟合对象,而非全部数据点,此时的离群点影响很低,因此几何体拟合的准确度得到了提高.



(a) 原点云图像



(b) 下采样后的点云图像

图3 5行5列下采样前后的点云图像对比

Fig. 3 Comparison between the point cloud images before and after downsampling with the step of 5 by 5

下一步对初步分割中每一个物体块进行下采样后的点云数据进行几何体拟合.图4是对墙壁的几何体拟合,其中:蓝色点为物体块的点云数据,红色立方体为几何体拟合后的结果.可以看出由于遮挡,墙壁的点云并不规则.进行几何体拟合后,墙壁的轮廓和姿态得到了改善.

图5是不同采样间隔下几何体拟合的时间曲线,其中:横坐标为不同的下采样间隔,纵坐标为每次几何体拟合的时间.可以看出,从不进行下采样到5行5列下采样的时间下降十分明显,而之后的时间改善并不显著.图6是不同采样间隔下拟合几何体的平均数量曲线,其中:横坐标为不同的下采样间隔,纵坐标为平均每帧拟合几何体的数量.可以看出,随着下采样间隔的增加,由于物体块轮廓丢失的情况越来越严重,一些物体块中的点数据将不足以进行几何体拟合,能够完成几何体拟合的物体块越来越少.综合这两方面,

本文选择5行5列为下采样间隔。

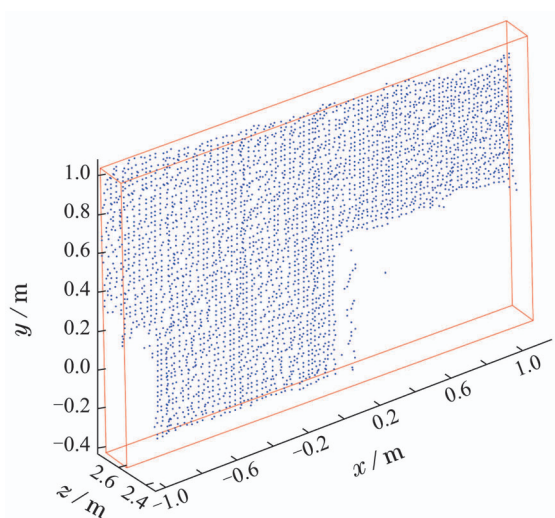


图 4 点云的几何体拟合

Fig. 4 Geometry fitting for points cloud

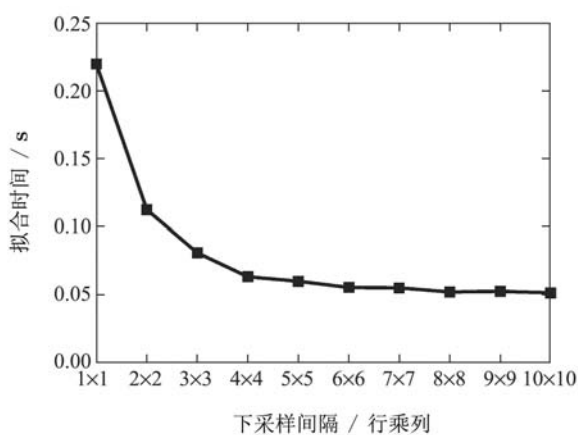


图 5 不同采样间隔下几何体拟合的时间曲线

Fig. 5 Time cost of geometry fitting with different downsampling steps

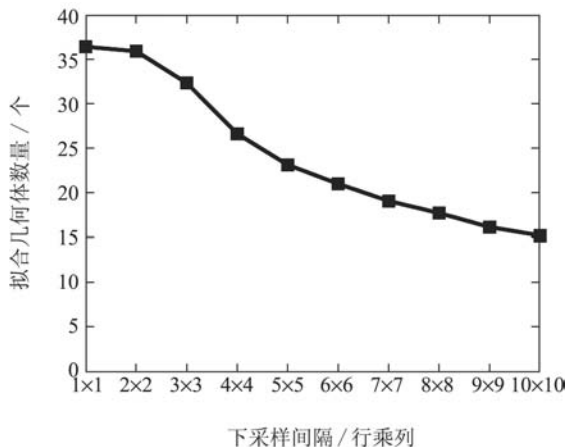


图 6 不同采样间隔下拟合几何体的平均数量曲线

Fig. 6 Average number of fitted geometry with different downsampling steps

4.2 基于支撑语义关系的图像分割实验结果

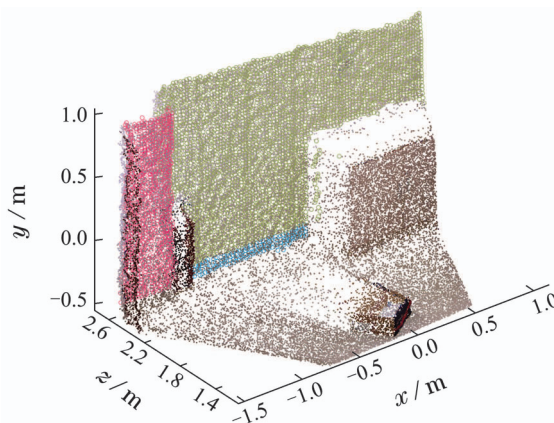
接下来根据物体块间的支撑语义概率以及单个物体的不稳定性函数和全局的能量函数, 使用SWC方法识别物体块间的强支撑语义关系, 并将具有强支撑语义关系的物体块进行合并, 形成完整物体。

支撑概率 q_e 与参数 ρ 正相关, 与 T_i, T_s, T_b 和 T_t 负相关, μ_g, μ_1 和 μ_2 为全局能量函数中不同组成成分的权重, μ_1 和 μ_2 越大, 本文算法所识别出的支撑语义关系越强, 物体块的分离程度越大. 实验中, 通过调节 μ_1 和 μ_2 来控制支撑语义关系的强弱以及物体的分离程度, 其他参数不变, 基本参数的设置为 $\rho = 1, T_i = 1, T_s = 1, T_b = 1, T_t = 1, \mu_g = 1, \mu_1 = 1$ 和 $\mu_2 = 1$. 实验平均每次迭代的用时为18.331 s, 平均每帧图像迭代次数为203.357次, 由于实验程序并未采用多线程也没有使用其它优化策略, 因此虽然算法复杂度不高但实验用时较长。

图7至图10展示了基于支撑语义关系的合并结果与先验知识不完备情况下的物体分割结果的直观对比. 在有色点云数据的基础上, 被标亮的数据点为本文算法所合并的物体块点云, 不同颜色表示了在先验知识不完备情况下的物体分割结果。



(a) 彩色图像



(b) 点云图像上的墙壁合并结果

图 7 基于支撑语义关系的合并结果1

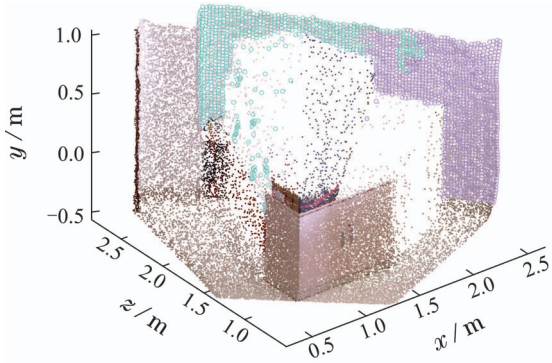
Fig. 7 Merging results 1 based on supportive semantic relationships



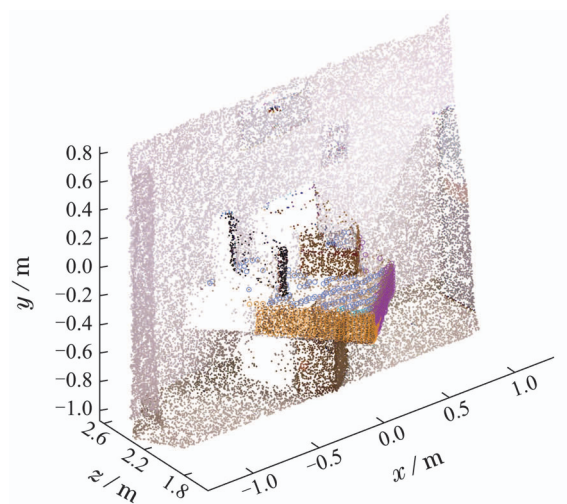
(a) 彩色图像



(a) 彩色图像



(b) 点云图像上的墙壁合并结果

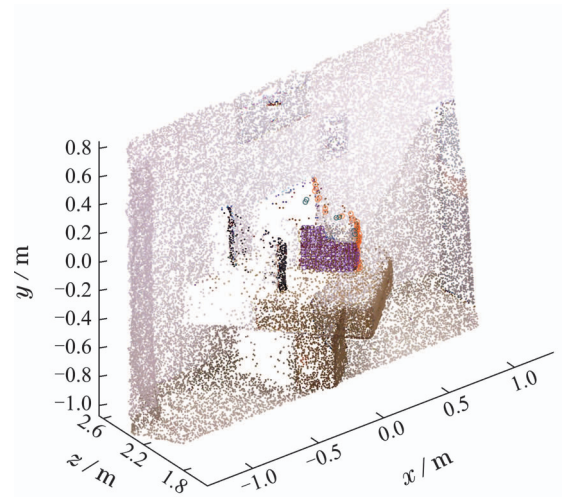


(b) 点云图像上的大箱子合并结果

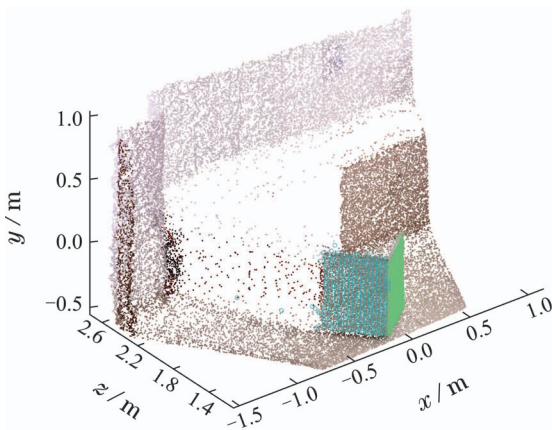
图8 基于支撑语义关系的合并结果2
Fig. 8 Merging results 2 based on supportive semantic relationships



(a) 彩色图像



(c) 点云图像上的小盒子合并结果



(b) 点云图像上的柜子合并结果

图9 基于支撑语义关系的合并结果3
Fig. 9 Merging results 3 based on supportive semantic relationships

图10 基于支撑语义关系的合并结果4
Fig. 10 Merging results 4 based on supportive semantic relationships

图7和图8中, 由于颜色和几何信息的不同与遮挡等原因, 造成完整的墙壁在先验知识不完备的情况下被过度分割成多个墙壁部分. 通过本文的方法, 墙壁部分之间的强支撑语义关系被识别出来, 因此将不同墙壁部分重新合并起来形成一个完整的墙壁物体.

图9中, 同一柜子的不同表面在先验知识不完备的

情况下被认为是不同物体,而在进行基于物体间支撑语义关系的图像分割后,算法识别出两个表面之间存在强支撑语义关系,因此将其合并为同一物体。

图10的场景中包含多个盒子和圆柱状的水壶。在先验知识不完备的情况下被分割开来的小盒子的2面与大箱子的3面均被成功地合并为同一个物体。表2给出了被过度分割的小盒子、大箱子以及水壶在合并前后的能量对比,参数为 $\mu_1 = 10^3$ 和 $\mu_2 = 10$ 。可以看出,由于大箱子部分之间与小盒子部分之间均存在强支撑语义关系,因此将它们分别进行合并都降低了能量函数的能量值。但是,由于大箱子和水壶之间存在弱支撑语义关系,合并大箱子和水壶反而增加了能量函数的能量值。本文算法成功地将强支撑语义关系和弱支撑语义关系区分开来。

表2 物体块合并前后的能量对比

Table 2 Energy difference of object parts before and after merging

合并情形	能量值
不合并	182.996
合并大箱子	177.914
合并小盒子	177.953
合并大箱子和水壶	177.961

表1给出了本文方法与初步分割的定量对比,虽然总准确率相同,但本文方法的精确率和召回率均高于初步分割方法。

由于本文对物体块合并的方法是无监督的,因此与其他无监督合并方法进行对比实验,包括亲和性传播 (affinity propagation, AP)算法^[21]和K-Means算法。其中,AP算法不需要指定类别个数,K-Means方法的类别数量由AP算法的结果得到,所使用的特征包括本文所给出的几何体属性以及RGB-D物体识别领域常见的特征^[22]。以初步分割后的物体块是否应该合并作为正例和负例,表3给出了精确率、召回率和准确率的对比结果。可以看出,虽然本文方法的召回率偏低,但精确率和准确率高于其他两种方法,这是由于聚类方法仅考虑了相似性,而不区分支撑语义关系的强弱,因此呈现过度合并的情况。

表3 本文方法与其他无监督方法的对比

Table 3 Comparison between other unsupervised methods and ours

方法	精确率/%	召回率/%	准确率/%
AP	29.59	4.25	76.58
K-Means	28.70	7.01	75.69
本文方法	50.02	1.59	77.86

实验结果表明,本文提出的基于物体间支撑语义关系的图像分割,能够在先验知识不完备的情况下,将同一物体被过度分割的部分重新合并起来,从而提升了图像分割的准确性。但是,本文的算法依赖于对参数的选择,参数选择影响了强、弱支撑语义关系的界限,进而影响了图像分割的质量。

5 结论

本文面向先验知识不完备的室内环境提出基于物体间支撑语义关系的RGB-D图像分割。针对先验知识不完备的过度分割结果,首先对每个物体块进行下采样和几何体拟合,然后根据统计得到的几何体属性计算物体块间相互存在支撑语义关系的概率,之后设计单个物体的不稳定性函数和全局的能量函数,利用SWC算法寻找物体块间的强支撑语义关系,最终根据强支撑语义关系合并物体完成图像分割。

本文的工作证明了引入物体间语义关系,特别是支撑语义关系,能够更好地实现针对室内环境的RGB-D图像分割,并克服先验知识不完备的情况。但是仍然存在一些被过度分割的物体部分没有被合并的情形,并且分割的质量依赖于参数选择,接下来将会继续完善支撑语义关系模型并引入参数的自适应选择,进一步提高图像分割的准确性和自主性。同时对算法进行优化,提高其可用性。此外,在未来将会研究更多种类的物体间语义关系并将其运用于其他场景理解问题当中。

参考文献:

- [1] GAO J, XIE Z. *Theory and Method of Image Understanding*. Beijing: Science Press, 2009.
- [2] GU G H. *Image semantic representation based scene classification research*. Beijing: Beijing Jiaotong University, 2013.
- [3] KHURANA P, SHARMA A, SINGH S N, et al. A survey on object recognition and segmentation techniques. *The 3rd International Conference on Computing for Sustainable Global Development*. Piscataway, USA: IEEE, 2016: 3822 – 3826.
- [4] MARTIN T. A survey of semantic segmentation. *Computing Research Repository*, 2016, abs/1602.06541: 1 – 16.
- [5] KHAN S H, HE X M, BANNAMOUN M, et al. Separating objects and clutter in indoor scenes. *IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, USA: IEEE, 2015: 4603 – 4611.
- [6] HAN F, ZHU S C. Bottom-up/top-down image parsing with attribute grammar. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(1): 59 – 73.
- [7] HANDA A, PATRAUCEAN V, STENT S, et al. SceneNet: an annotated model generator for indoor scene understanding. *IEEE International Conference on Robotics and Automation*. Piscataway, USA: IEEE, 2016: 5737 – 5743.
- [8] VENKATESHKUMAR S K, SRIDHAR M, OTT P. Latent hierarchical part based models for road scene understanding. *IEEE International Conference on Computer Vision Workshop*. Piscataway, USA: IEEE, 2015: 115 – 123.

- [9] KUNZE L, BURBRIDGE C, ALBERTI M, et al. Combining top-down spatial reasoning and bottom-up object class recognition for scene understanding. *IEEE/RSJ International Conference on Intelligent Robots and Systems*. Piscataway, USA: IEEE, 2014: 2910 – 2915.
- [10] ZHENG B, ZHAO Y, YU J C, et al. Beyond point clouds: scene understanding by reasoning geometry and physics. *IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, USA: IEEE, 2013: 3127 – 3134.
- [11] JIA Z, GALLAGHER A C, SAXENA A, et al. 3D reasoning from blocks to stability. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(5): 905 – 918.
- [12] ZHENG B, ZHAO Y, YU J C, et al. Detecting potential falling objects by inferring human action and natural disturbance. *IEEE International Conference on Robotics and Automation*. Piscataway, USA: IEEE, 2014: 3417 – 3424.
- [13] DUPRE R, ARGYRIOU V. 3D voxel HOG and risk estimation. *IEEE International Conference on Digital Signal Processing*. Piscataway, USA: IEEE, 2015: 482 – 486.
- [14] DUPRE R, ARGYRIOU V, GREENHILL D, et al. A 3D scene analysis framework and descriptors for risk evaluation. *International Conference on 3D Vision*. Piscataway, USA: IEEE, 2015: 100 – 108.
- [15] WANG Xiaonian, FENG Yuanjing, FENG Zuren. Ant colony optimization with active contour models for image segmentation. *Control Theory & Applications*, 2006, 23(4): 515 – 522.
(王晓年, 冯远静, 冯祖仁. 一种基于主动轮廓模型的蚁群图像分割算法. *控制理论与应用*, 2006, 23(4): 515 – 522.)
- [16] SILBERMAN N, HOIEM D, KOHLI P, et al. Indoor segmentation and support inference from RGBD images. *European Conference on Computer Vision*. Berlin, Germany: Springer, 2012: 746 – 760.
- [17] FISCHLER M A, BOLLES R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 1981, 24(6): 381 – 395.
- [18] ZHU S C, BARBU A. Generalizing Swendsen-Wang to sampling arbitrary posterior probabilities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(8): 1239 – 1253.
- [19] BARBU A, ZHU S C. Graph partition by Swendsen-Wang cuts. *IEEE International Conference on Computer Vision*. Piscataway, USA: IEEE, 2003: 320 – 327.
- [20] BARBU A, ZHU S C. Multigrid and multi-level Swendsen-Wang cuts for hierarchic graph partition. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Piscataway, USA: IEEE, 2004: 731 – 738.
- [21] DUECK D. *Affinity propagation: clustering data by passing messages*. Canadian: University of Toronto, 2009.
- [22] RUIZ-SARMIENTO J R, GALINDO C, GONZALEZ-JIMENEZ J. A survey on learning approaches for undirected graphical models. Application to scene object recognition. *International Journal of Approximate Reasoning*, 2017, 83(1): 434 – 451.

作者简介:

陈昊天 博士研究生, 目前研究方向为智能机器人系统, E-mail: archristy@hotmail.com;

郑阳 硕士研究生, 目前研究方向为智能机器人系统, E-mail: 584846936@qq.com;

张钰桐 硕士研究生, 目前研究方向为智能机器人系统, E-mail: 1969536200@qq.com;

孙凤池 副教授, 硕士生导师, 目前研究方向为机器学习、智能机器人系统, E-mail: fengchisun@nankai.edu.cn;

黄亚楼 教授, 博士生导师, 目前研究方向为智能机器人系统、数据挖掘, E-mail: huangyl@nankai.edu.cn.