Journal of System Simulation

Volume 34 | Issue 2 Article 10

2-23-2022

Monocular Semantic SLAM Method Based on Object Relation Description

Shiqi Lin
Department of Automation, University of Science and Technology of China, Hefei 230027, China;,
linshiqi@mail.ustc.edu.cn

Jikai Wang
Department of Automation, University of Science and Technology of China, Hefei 230027, China;

Haoyuan Pei
Department of Automation, University of Science and Technology of China, Hefei 230027, China;

Hao Zhao
Department of Automation, University of Science and Technology of China, Hefei 230027, China;

See next page for additional authors

Monocular Semantic SLAM Method Based on Object Relation Description

Abstract

Abstract: Semantic information perception of the external environment and accurate positioning are the keys to autonomous navigation and operation of mobile robots. This paper proposes a method of semantic simultaneous localization and mapping (SLAM) based on a monocular camera. The system completes three-dimensional (3D) object detection while estimating the trajectory. *We model the 3D objects with cuboids. Then, the semantic meanings, color distribution, size and neighborhood topology of the objects are extracted as descriptors for the accurate matching of objects between different frames. The camera pose, map points and object landmarks are optimized jointly in the backend of the system. The weight coefficient of each error term in the cost function is autonomously adjusted to improve the estimation accuracy and robustness of each state variable of the system.* The experimental results show that the proposed method has high accuracy in map construction.

Keywords

image segmentation, 3D object detection, topological map, graph matching, semantic SLAM

Authors

Shiqi Lin, Jikai Wang, Haoyuan Pei, Hao Zhao, and Zonghai Chen

Recommended Citation

Shiqi Lin, Jikai Wang, Haoyuan Pei, Hao Zhao, Zonghai Chen. Monocular Semantic SLAM Method Based on Object Relation Description[J]. Journal of System Simulation, 2022, 34(2): 278-284.

第 34 卷第 2 期 系统仿真学报© Vol. 34 No. 2
2022 年 2 月 Journal of System Simulation Feb. 2022

基于物体关系描述的单目语义SLAM方法

林士琪，王纪凯，裴浩渊，赵皓，陈宗海
(中国科学技术大学 自动化系，安徽 合肥 230027)

摘要：外界环境的语义感知和自身位置的准确估计是移动机器人自主导航和作业的关键。提出了一种基于单目相机的语义SLAM（simultaneous localization and mapping）方法，*在轨迹估计的同时完成三维目标检测。提取物体自身语义、尺寸、颜色分布及其邻域拓扑结构等多元信息作为描述子，实现帧间物体的准确关联。在后端对相机位姿、地图点和物体路标进行联合优化，并自适应调整代价函数中各误差项的权重系数，以提高各状态变量的估计精度和鲁棒性。*实验结果表明，所提出的算法在地图构建方面具有较高的精度。

关键词：图像分割；三维目标检测；拓扑地图；图匹配；语义SLAM
中图分类号：TP391 文献标志码：A 文章编号：1004-731X(2022)02-0278-07
DOI: 10.16182/j.issn1004731x.joss.20-0734

Monocular Semantic SLAM Method Based on Object Relation Description

Lin Shiqi, *Wang Jikai*, *Pei Haoyuan*, *Zhao Hao*, *Chen Zonghai*
(Department of Automation, University of Science and Technology of China, Hefei 230027, China)

Abstract: Semantic information perception of the external environment and accurate positioning are the keys to autonomous navigation and operation of mobile robots. This paper proposes a method of semantic simultaneous localization and mapping (SLAM) based on a monocular camera. The system completes three-dimensional (3D) object detection while estimating the trajectory. *We model the 3D objects with cuboids. Then, the semantic meanings, color distribution, size and neighborhood topology of the objects are extracted as descriptors for the accurate matching of objects between different frames. The camera pose, map points and object landmarks are optimized jointly in the backend of the system. The weight coefficient of each error term in the cost function is autonomously adjusted to improve the estimation accuracy and robustness of each state variable of the system*. The experimental results show that the proposed method has high accuracy in map construction.

Keywords: image segmentation; 3D object detection; topological map; graph matching; semantic SLAM

引言

基于语义信息的自主定位和建图是目前的研究热点之一，语义信息能够提供丰富的环境信息和观测约束，有利于机器人对环境的感知与认知。语义感知和SLAM是一个相互促进的过程。SLAM算法在运行过程中可以实时提供相机位姿，并在运动过程中对同一目标可产生多次观测，这将有利于进行目标检测任务。文献[1-3]利用上述优势，通过SLAM辅助目标检测，实现了检测精度的提升。点特征在轨迹估计的过程中仅能为相机位姿提供短期约束，在大尺度空间范围中容易产生较大偏差。相比之下，物体作为更高层级的特征，其语义信息不会因光照

收稿日期：2020-09-23 修回日期：2020-12-10
基金项目：国家自然科学基金(91848111，61703387)
第一作者：林士琪(1992-)，男，博士生，研究方向为视觉SLAM。E-mail：linshiqi@mail.ustc.edu.cn

强度、观测距离和角度的变化而发生改变，在环境中更加稳定，可以为相机位姿提供较长时间的约束。文献[2, 4-5]利用这一良好特性，将环境中存在的高层语义特征加以利用，实现了定位精度的提升。

在语义地图构建方面，很久之前便有学者开始了相关研究。文献[6]提出了一种基于物体的SLAM方法，与传统基于点、线、面等低层级特征基元的SLAM方法不同，系统通过将RGB-D相机采集的数据与预先建好的3D模型数据库进行匹配，实现对环境中存在的椅子、桌子等物体进行检测和定位并在地图中进行显示。但由于需要事先构建物体模型，导致系统通用性较差。文献[7]首次提出基于二次曲线的物体表示方法，之后被文献[8]应用在SFM(structure from motion)中。文献[9]采用基于RGB-D相机的SLAM与目标检测相结合的方法为环境中的物体重建出准确的模型，然而所构建的物体模型没有用在轨迹估计中。文献[10]提出了著名的CubeSLAM开源框架，算法前端利用基于消失点采样的方法从单张图片中检测出三维物体，并定义基于物体的重投影误差，将其放入后端中进行非线性优化，通过连续多帧观测实现稳定的三维目标检测。本文将以此开源框架为基础，结合物体自身语义、尺寸、颜色分布及其邻域拓扑结构等多元信息作为描述子，用于物体间的关联匹配。在后端对相机位姿、地图点和物体路标进行联合优化，并根据二维目标检测质量和特征点属性自适应调整代价函数中3种误差项的权重系数，实现系统各状态量的准确估计。

1 算法设计

本文提出的基于单目的语义SLAM框架如图1所示，系统首先通过单目图像对当前相机的位姿进行估计。与此同时，利用YOLOv3[11]对当前输入图像进行目标检测，并采用CubeSLAM提出的方法，结合当前相机位姿以及二维目标检测结果，实现基于单张图像的三维目标检测。在帧间位姿估计、地图构建、相机重定位和闭环检测等过程中物体间的准确关联匹配至关重要。本文抽取其语义信息、颜色信息、尺寸信息以及邻域拓扑信息作为物体的唯一性描述方式，以确保关联匹配的准确性。然后，基于捆集调整，对相机位姿、三维目标和地图点等变量进行联合批量优化，实现噪声中各状态量的准确估计。最后，通过闭环检测消除累积误差，得到全局一致的相机轨迹和语义地图。

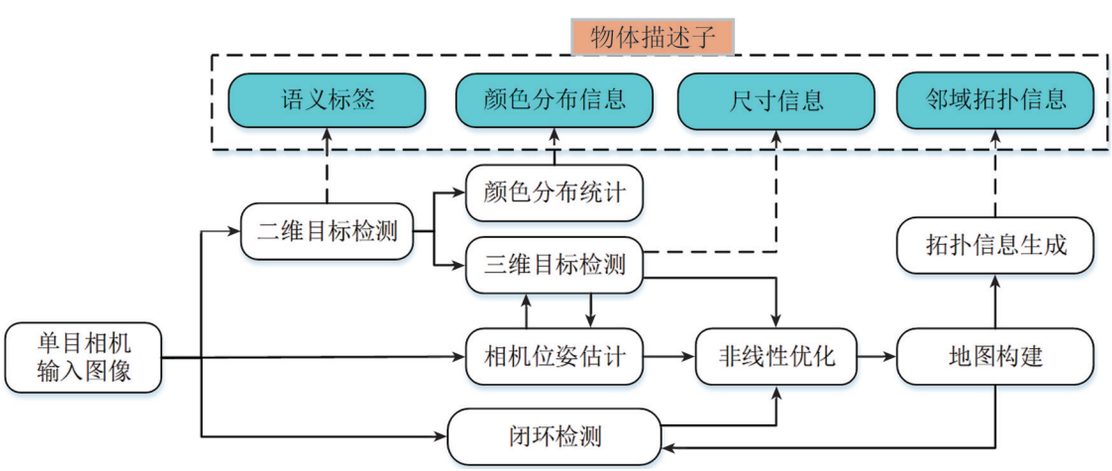


图1 系统框架
Fig. 1 System framework

DOI: 10.16182/j.issn1004731x.joss.20-0734

1.1 色彩分布信息提取

环境中物体的外表色彩分布情况是其外观特异性的重要表示形式之一，本文中将提取此类信息并将其作为描述子的一部分用于物体间的关联匹配。

色彩分布信息的提取过程如图2所示，首先采用YOLOv3[11]对当前输入图像进行物体检测和定位。通常矩形框中包含的背景颜色大致相同，并且与物体颜色相差较大，本文基于这一先验信息，采用K-means算法对像素进行聚类，将其去除。主流色彩空间分为RGB空间和HSV空间2种，实验结果表明，2种色彩空间对不同颜色物体的聚类效果各有优劣。鉴于此类特性，本文提出自适应选择策略，首先对输入图像在2种空间中分别进行聚类，然后自适应选择较好的一种进行后续的背景去除工作。由于采用颜色直方图进行物体关联的方法具有较高的鲁棒性，对背景去除操作具有一定的容错能力，在少量像素识别错误的情况下关联匹配仍不会出错。

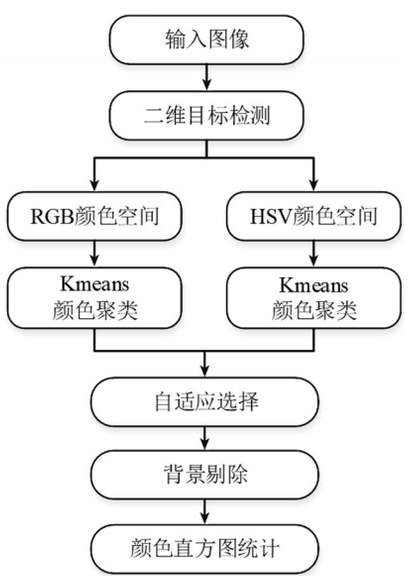


图2 颜色分布信息提取

Fig. 2 Extraction of color distribution information

1.2 三维目标检测和尺寸信息提取

本文采用CubeSLAM中提出的方法实现三维目标检测任务。首先利用YOLOv3对二维图像进行目标检测，然后利用消失点推测出长方体的八个顶点在二维图像中的坐标，通过当前相机位姿将八个顶点投影到三维空间中，实现基于单张图片的三维目标检测。最后将其加入后端非线性优化框架中，通过连续多帧观测产生的约束进一步提升三维目标检测质量。

长、宽、高等信息作为物体外观尺寸的表达，不随观测距离和观测角度的变化而发生变化，具有相对稳定的性质。本文将在系统运行过程中对此类信息进行实时检测和提取，一并纳入物体描述子中。对于环境中存在的物体，通过上述提到的三维目标检测方法可以得到包围该物体的3D长方体，进而得到物体的尺寸信息。

1.3 拓扑信息提取

当环境中存在过多的相同物体时，单纯依靠自身语义、颜色和尺寸等信息无法对其进行唯一性描述，可能导致物体和场景关联失败，这种情况在闭环检测环节中时有发生。本文深度挖掘环境中潜在的拓扑信息，通过对物体周围的拓扑结构进行编码，统计邻域内其他物体的分布情况及其相对位置关系，在关联匹配过程中同时核验此部分信息，以消除歧义性，达到准确关联的目的。

本文首先将环境中存在的物体抽象为节点，并将以该物体为中心、一定长度为半径的球形区域内所涉及的其他物体用边连接起来，边的权重设置为两物体间的距离，由此可形成一个表征环境中物体间复杂相互关系的无向连接图。在对某一节点进行编码时，以该节点为起点，在无向连接图中进行随机游走[12-13]，每次游走的深度以及游走的次数视场景分布情况而定，游走过程中记录经过的节点以及边的权重，如图3所示。通过上述方式得到了可以表示物体周围拓扑关系的矩阵，称为拓扑描述子，通过描述子匹配可以实现场景拓扑结构的相似性度量。

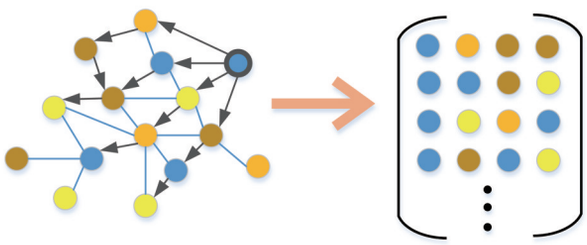


图3 拓扑描述子提取
Fig. 3 Extraction of topology descriptor

1.4 非线性优化

系统前端视觉里程计可以在短时间内对相机位姿和地图点做出估计，但由于存在不可避免的观测噪声，上述量的估计是不准确的。此外，由于缺乏深度信息，基于单张图片的三维目标检测方法具有较大不确定性，往往准确性得不到保证。针对上述存在的问题，本文采用批量估计的方法，同时使用过去和未来的信息来更新当前的状态，对一段时间窗口内的相机位姿、地图点和三维目标等状态量进行联合优化，得到更加准确的系统状态估计结果。

将相机、三维目标和地图点分别表示为 $C=\{C_i\}$，$O=\{O_j\}$，$P=\{P_k\}$。其中，相机位姿包含 6 个自由度，表示为 $\boldsymbol{T}_c\in SE(3)$。地图点坐标有 3 个自由度，将其表示为 $\boldsymbol{P}\in\mathbb{R}^3$。为物体建模的立方体由 $O=\{\boldsymbol{T}_o,\boldsymbol{d}\}$ 表示，其中 $\boldsymbol{T}_o=[\boldsymbol{R},\boldsymbol{t}]\in SE(3)$ 存储平移和旋转信息，$\boldsymbol{d}=\{dx,dy,dz\}\in\mathbb{R}^3$ 存储尺度信息。BA 可以归结为如式(1)所示的优化求解问题：

$$\{\hat{C}\},\{\hat{O}\},\{\hat{P}\}=\underset{\{C,O,P\}}{\arg\min}\,E_{cp}+\lambda_1E_{co}+\lambda_2E_{op} \tag{1}$$

式中：E_{cp},E_{co},E_{op} 分别代表相机与特征点之间、相机与物体之间以及物体与特征点之间的投影误差。λ_1,λ_2 分别为 2 个误差项的权重系数。此类问题可以通过 Levenberg–Marquardt 或 Gauss–Newton 等算法迭代优化求解，具体可以调用 g2o[14]实现。下面将对上述 3 种测量误差以及权重系数自适应调整策略一一进行详细论述。

1.4.1 相机与特征点间的误差

将三维地图点投影到二维图像平面，测量其与对应特征点间的像素距离，即为特征点与相机之间的重投影误差：

$$E_{cp}=\sum_i\sum_k\left\|e_{cp}(i,k)\right\|^2_{\Sigma_{ik}} \tag{2}$$

$$e_{cp}=\pi\left(\boldsymbol{T}_c^{-1}\boldsymbol{P}\right)-\boldsymbol{z}_m \tag{3}$$

1.4.2 物体与相机间的误差

物体与相机之间的误差项由两部分组成，分别为三维空间中的测量误差 e_{co_3D} 和二维平面中的测量误差 e_{co_2D}。

$$E_{co}=\sum_i\sum_j\left\|e_{co\_3D}(i,j)\right\|^2_{\Sigma_{ij}}+\left\|e_{co\_2D}(i,j)\right\|^2_{\Sigma_{ij}} \tag{4}$$

3D 测量误差：通过消失点采样法从单张图片中生成的物体位姿为 $O_m=\left(\boldsymbol{T}_{om},\boldsymbol{d}_m\right)$，假设地图中存在与其相对应的、准确的 3D 立方体路标 $O=\{\boldsymbol{T}_o,\boldsymbol{d}\}$，将路标转换到相机坐标系下，与 $O_m=\left(\boldsymbol{T}_{om},\boldsymbol{d}_m\right)$ 进行比较得到测量误差：

$$e_{co\_3D}=\begin{bmatrix}\log\left(\left(\boldsymbol{T}_c^{-1}\boldsymbol{T}_o\right)\boldsymbol{T}_{om}^{-1}\right)^{\vee}_{se3} & \boldsymbol{d}-\boldsymbol{d}_m\end{bmatrix} \tag{5}$$

其中，log 操作将 $SE3$ 类型的误差映射到李代数形式，因此 $e_{co\_3D}\in\mathbb{R}^9$。

2D 重投影误差：首先将地图中长方体的八个顶点坐标按照当前相机位姿投影到图像平面，并依次求取八个顶点在像素坐标系下的最小和最大坐标值：

$$\begin{aligned}[u,v]_{\min}&=\min\left\{\pi\left(\boldsymbol{R}\left[\pm\boldsymbol{d}_x,\pm\boldsymbol{d}_y,\pm\boldsymbol{d}_z\right]/2+\boldsymbol{t}\right)\right\}\\[u,v]_{\max}&=\max\left\{\pi\left(\boldsymbol{R}\left[\pm\boldsymbol{d}_x,\pm\boldsymbol{d}_y,\pm\boldsymbol{d}_z\right]/2+\boldsymbol{t}\right)\right\}\end{aligned} \tag{6}$$

其中，π 操作代表将三维坐标点投影到二维平面。计算矩形框的中心点坐标和尺寸信息，$\boldsymbol{c}\in\mathbb{R}^2$ 代表矩形框的中心点坐标，$\boldsymbol{s}\in\mathbb{R}^2$ 代表其尺寸信息。

$$\begin{aligned}\boldsymbol{c}&=\left([u,v]_{\min}+[u,v]_{\max}\right)/2\\\boldsymbol{s}&=[u,v]_{\max}-[u,v]_{\min}\end{aligned} \tag{7}$$

通过比较 2 个矩形框的上述参数得到测量误差：

https://dc-china-simulation.researchcommons.org/journal/vol34/iss2/10
DOI: 10.16182/j.issn1004731x.joss.20-0734

$$e_{co\_2D}=\left[\boldsymbol{c},\boldsymbol{s}\right]-\left[\boldsymbol{c}_m,\boldsymbol{s}_m\right] \tag{8}$$

由于3D长方体投影后将会丢失很多信息，3D长方体与2D矩形框之间存在着一对多的关系，因此靠少量图像帧中的2D矩形框很难对3D长方体进行较强的约束，需要连续观测多帧才能产生较好的结果。

1.4.3 物体与特征点间的误差

如果在二维图像中特征点属于某一物体，那么在三维空间中，其对应的地图点应该位于长方体内部，本文将利用此信息生成物体与特征点之间的误差函数。当地图点位于长方体内部时，令误差为零。地图点处于长方体外部时，测量地图点与长方体的距离，将其作为物体与特征点间的误差：

$$E_{op}=\sum_j\sum_k\left\|e_{op}(j,k)\right\|^2_{\sum_{jk}} \tag{9}$$

$$e_{op}=\max\left(\left|\boldsymbol{T}_o^{-1}\boldsymbol{P}\right|-\boldsymbol{d}_m,\boldsymbol{0}\right) \tag{10}$$

1.4.4 误差权重自适应调整策略

BA的误差代价函数由3项组成，在优化过程中需要通过两个权重系数λ_1,λ_2分别调节物体与相机和物体与地图点间两个误差项在总的代价函数中所占的比重。本文采用自适应调整策略，根据物体检测的质量以及特征点的属性动态调整λ_1,λ_2，以提升BA优化的效果。下面将对两个权重系数的调整策略一一进行详细介绍。

以目标检测网络输出的矩形框作为基准，将地图中3D路标投影到当前相机平面与其进行比较，进而得到重投影误差。由此过程可以看出，此误差项假设目标检测网络可以输出精准的矩形框，并以此为前提进行误差具体值的计算并对其进行优化。当目标检测输出的矩形框存在误差时，以它为基准计算出的重投影误差项往往较大，这会导致BA优化过程中算法重点优化此误差项，过度调整相机位姿和物体路标使得误差项变小，从而输出错误的优化结果。针对上述存在的问题，本文以目标检测输出置信度值为参考，动态调整权重系数的大小。置信度越大，说明矩形框检测质量高，其权重系数λ_1也应随之增大。λ_1与矩形框的置信度b_{conf}间的数学关系如下：

$$\lambda_1=k_1\cdot\exp\left(\frac{{b_{\text{conf}}}^2}{{\sigma_b}^2}\right) \tag{11}$$

在计算物体与特征点间的误差时，需要确定二维图像上属于该物体的特征点在三维空间中是否处于对应的3D长方体内部。出于系统效率的考虑，本文没有采用基于深度学习的语义分割方案，而利用颜色空间聚类的方法进行前景图像分割，此方法无法准确的为每个像素点确定类别归属。当处于背景中的特征点被误认为属于物体时，便出现一个较大的误差项，从而会对优化过程造成干扰。针对上述存在的问题，本文同时从空间距离和灰度距离两个角度为特征点与物体间的距离进行度量，并将此作为该特征点是否属于物体的一种概率度量，由此定义误差权重。空间距离和灰度距离由等式（12）和等式（13）定义：

$$d_{spa}(x,y)=\exp\left(\frac{\left(x_u-y_u\right)^2+\left(x_v-y_v\right)^2}{2\sigma_d^2}\right) \tag{12}$$

$$d_{gray}(x,y)=\exp\left(\frac{\left(gray\left(x_u,x_v\right)-gray\left(y_u,y_v\right)\right)^2}{2\sigma_g^2}\right) \tag{13}$$

式中：$\left(x_u,x_v\right)$为特征点坐标；$\left(y_u,y_v\right)$为物体中心点坐标。权重系数λ_2与特征点和物体间的距离成反比：

$$\lambda_2=\frac{k_2}{d_{\text{spa}}\cdot d_{\text{gray}}} \tag{14}$$

2 实验与分析

本节从物体背景去除效果、语义地图生成质量以及物体建模精度等方面对提出的算法进行检验，并与CubeSLAM算法进行对比分析。

2.1 物体背景去除

图像背景的去除效果直接影响下一步对物体

自身颜色分布信息的提取的准确度，因此本文基于自己采集的数据对背景去除算法进行评估。实验结果如图4所示，从图中可以看到在HSV颜色空间中对红色物体聚类效果较好，对绿色物体聚类效果较差。而在RGB颜色空间正好相反。这说明在两种空间中各种颜色间的距离相差较大。图中最后一列是最终背景去除的效果，可见算法可以较好的提取出仅包含物体的前景图像，满足实际需求。

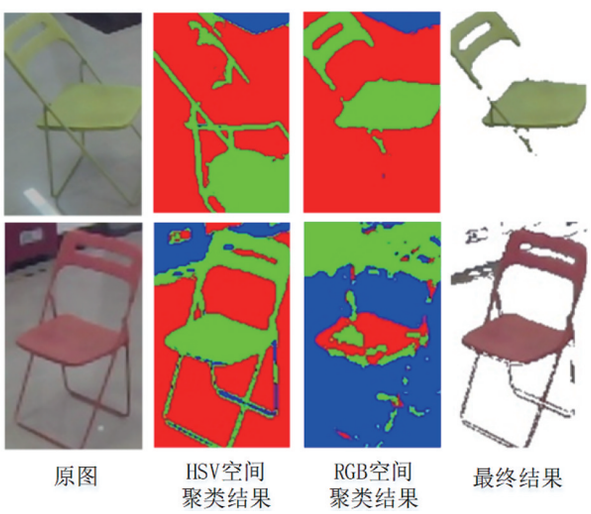
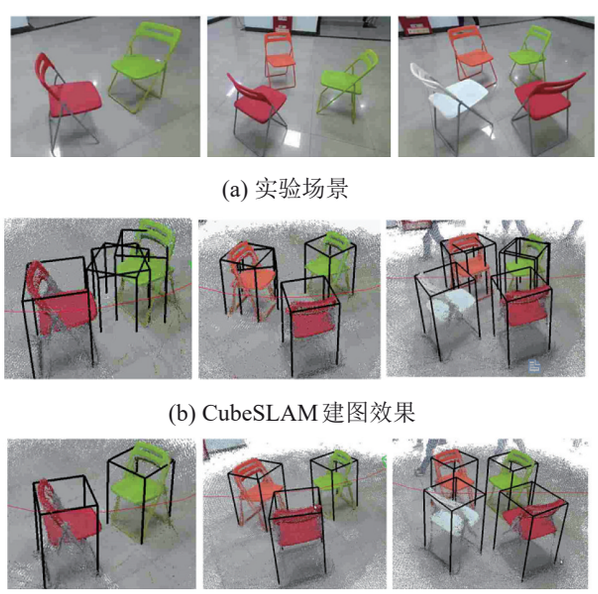


图4　背景去除
Fig. 4　Background removal

2.2 语义地图构建

本文在环境中存放若干把椅子，令相机围绕其周围一圈采集数据，利用CubeSLAM和本文提出的算法分别进行了实验，检测语义地图生成质量。

本文所提出的方法将单目相机所采集的图片序列作为输入，实时输出每帧相机位姿以及由稀疏地图点和三维长方体组成的环境语义地图。为了对语义地图的构建精度进行可视化，采用RGB-D相机采集环境数据，对当前实验场景进行稠密重建，同时将其中的RGB图作为输入构建语义地图，并将两者构建的稠密点云地图和语义地图进行融合显示，建图效果如图5所示，其中黑色长方体代表语义地图中的三维物体模型。从图中可以看出，CubeSLAM算法构建的语义地图出现错误，多检测出了若干个物体。这主要是因为基于共享地图点数量的物体关联方法鲁棒性较低，在大视角观测以及物体纹理简单导致的共享地图点过少的情况下会造成关联失败，导致系统误以为是首次检测到此物体，而再次为其重新构建一个模型。从第三行语义地图构建效果中可以发现，本文提出的方法较为准确，没有出现误检测的情况。这充分说明了本文所提出的基于物体描述子的物体关联方法不受物体自身纹理以及观测视角约束的限制，具有较高的鲁棒性。



(a) 实验场景

(b) CubeSLAM建图效果

(c) 本文方法建图效果

图5　语义建图质量对比
Fig. 5　Semantic mapping quality comparison

此外，以重建的稠密点云地图为基础，对环境中存在的物体进行手动标注，并计算所标注立方体与相对应的算法检测的立方体之间的交并比(3D IoU)，IoU数值越大，代表物体建模精度越高。本文将其作为指标，针对两种方法的物体建模精度进行了量化对比评估，如表1所示。从表中可以看出，本文提出的方法可以有效提高物体建模精度，改善语义地图构建质量。

https://dc-china-simulation.researchcommons.org/journal/vol34/iss2/10
DOI: 10.16182/j.issn1004731x.joss.20-0734

第 34 卷第 2 期 系统仿真学报 Vol. 34 No. 2
2022 年 2 月 Journal of System Simulation Feb. 2022

表 1 物体检测精度（IoU）对比
Table 1 Comparison of IoU on benchmark

算法	01	02	03	04	05	06
CubeSLAM	0.57	0.33	0.62	0.61	0.50	0.56
Ours	**0.63**	**0.62**	**0.69**	**0.72**	**0.58**	**0.73**

3 结论

本文提出了一个基于单目相机的新型语义SLAM方法，在位姿估计的同时对环境中的物体进行三维感知。针对基于共有地图点数量的物体关联方法存在易受物体表面纹理以及观测角度影响的问题，本文提出了基于语义、尺寸、颜色分布以及邻域拓扑结构等多元信息的物体描述方法，以保障物体关联匹配的准确性。此外，针对语义SLAM系统，提出了一种更加鲁棒的BA优化目标函数，以充分利用观测之间的约束信息实现准确的状态估计。经实验验证，本文提出的算法在语义地图构建质量方面优于目前主流同类型算法。下一步，将继续挖掘物体建模在语义SLAM中的应用潜力，将所构建的物体模型以及物体描述子应用在闭环检测和相机重定位任务中，进一步提高系统的估计精度和稳定性。

参考文献

[1] Dong J, Fei X, Soatto S. Visual-inertial-semantic Scene Representation for 3D Object Detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 960-970.

[2] Gálvez-López D, Salas M, Tardós J D, et al. Real-time Monocular Object Slam[J]. Robotics and Autonomous Systems(S0921-8890), 2016, 75: 435-449.

[3] Leibe B, Cornelis N, Cornelis K, et al. Dynamic 3d Scene Analysis from a Moving Vehicle[C]//2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2007: 1-8.

[4] Gay P, Rubino C, Bansal V, et al. Probabilistic Structure from Motion with Objects (Psfmo) [C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 3075-3084.

[5] Lianos K N, Schonberger J L, Pollefeys M, et al. Vso: Visual Semantic Odometry[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 234-250.

[6] Salas-Moreno R F, Newcombe R A, Strasdat H, et al. Slam++: Simultaneous Localisation and Mapping at the Level of Objects[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013: 1352-1359.

[7] Cross G, Zisserman A. Quadric reconstruction from Dual-space Geometry[C]//Sixth International Conference on Computer Vision IEEE Cat. No. 98CH36271. IEEE, 1998: 25-31.

[8] Sünderhauf N, Pham T T, Latif Y, et al. Meaningful Maps with Object-Oriented Semantic Mapping[C]//2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2017: 5079-5085.

[9] Sünderhauf N, Pham T T, Latif Y, et al. Meaningful Maps with Object-Oriented Semantic Mapping[C]//2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2017: 5079-5085.

[10] Yang S, Scherer S. Cubeslam: Monocular 3-d Object Slam[J]. IEEE Transactions on Robotics(S1941-0468), 2019, 35(4): 925-938.

[11] Redmon J, Farhadi A. Yolov3: An Incremental Improvement[C]// 2018 IEEE International Conference on Computer Vision and Pattern Recognition, IEEE, 2018: 1-6.

[12] Gawel A, Del Don C, Siegwart R, et al. X-view: Graph-based Semantic Multi-View Localization[J]. IEEE Robotics and Automation Letters(S2377-3766), 2018, 3(3): 1687-1694.

[13] Liu Y, Petillot Y, Lane D, et al. Global Localization with Object-Level Semantics and Topology[C]// 2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019: 4909-4915.

[14] Kümmerle R, Grisetti G, Strasdat H, et al. g 2 o: A General Framework for Graph Optimization[C]//2011 IEEE International Conference on Robotics and Automation. IEEE, 2011: 3607-3613.

