

激光与视觉融合识别并构建语义地图改善定位性能

蒋林^{1,2}, 刘奇^{1*}, 雷斌^{1,2}, 左建朋¹, 赵慧^{2,3}¹ 武汉科技大学冶金装备及其控制教育部重点实验室, 湖北 武汉 430081;² 武汉科技大学机器人与智能系统研究院, 湖北 武汉 430081;³ 武汉科技大学机械传动与制造工程湖北省重点实验室, 湖北 武汉 430081

摘要 语义地图能够帮助移动机器人更好地理解环境,改善定位性能。为了提高语义地图的精度,本文提出一种激光与视觉融合识别并构建语义地图的方法。首先对激光雷达击中点进行直线拟合,提取凸凹墙角信息;然后利用相机和电子罗盘进行墙角的方向性进行区分,得到丰富的墙角类别信息;接着利用墙角信息对柜式物体进行误识别去除,融合激光深度信息,得到更加精确的物体语义信息;最后利用贝叶斯估计增量式地构建语义地图,消除语义栅格重叠构建的问题。通过仿真环境以及真实环境下的实验验证了该方法的有效性,同时利用语义地图进行重定位实验验证了该语义地图能有效改善定位性能。

关键词 遥感; 直线拟合; 墙角类别; 贝叶斯估计; 语义地图; 重定位

中图分类号 TP249

文献标志码 A

DOI: 10.3788/CJL202249.1810003

1 引言

移动机器人在进行位姿估计时需要知道环境地图信息,而创建环境地图时又需要精确的定位,因此,同时定位与建图(SLAM)就将两者合二为一,在创建地图的同时进行位姿估计^[1-2]。传统 SLAM 方法构建的环境地图只能表达环境中的拓扑信息与几何信息^[3],缺少对环境语义信息的获取与描述,导致机器人无法真正地理解环境。语义地图可以丰富地图信息,帮助机器人实现更加智能的定位^[4-5]与导航^[6]。

SLAM 算法是语义地图构建的基础,语义地图的构建建立在 SLAM 算法之上,且不同于 SLAM 算法^[7]。语义地图在构建过程中不仅要进行基本的 SLAM 建图,还要对环境中的独立物体进行识别,获取其位置和轮廓等语义信息,并将这些信息融入地图中。

文献[8]提出了一种基于 DA-RNN 网络的语义地图构建方法,该方法将网络的输出结果与 KinectFusion^[9]的 SLAM 技术相结合,通过将检测到的语义信息注入到三维重建地图中构建语义地图。KinectFusion 是最早的基于 RGB-D 的实时三维重建算法,之后出现的 ElasticFusion^[10]算法的三维重建效果更好。文献[11]基于 ElasticFusion 和卷积神经网络(CNN)提出了一种稠密的三维语义映射算法 SemanticFusion。文献[12]提出了基于 LSD-SLAM

和卷积神经网络的算法,并采用该算法构建了半稠密语义地图。该算法与 SemanticFusion 类似,首先通过 LSD-SLAM 估计相机位姿,构建半稠密的环境地图,然后使用 DeepLab-v2 进行语义分割,获取语义信息,最后利用贝叶斯进行数据关联,构建半稠密语义地图。文献[13]对三维激光进行语义分割后提取目标语义信息。文献[14]基于 ORB-SLAM2 构建稠密点云地图,并将三维点云分割成类别明显的区段,同时利用 YOLOv3 算法提取语义信息,融合点云标签,构建稠密的点云地图。

视觉传感器本身容易受到环境影响而形成累积误差,导致地图的鲁棒性不高。激光雷达相较于相机来说,精度更高,受环境的影响更小。因此,很多学者基于激光 SLAM 建图,并融合视觉的语义信息来构建语义地图。文献[15]使用 2D 激光雷达和 Gmapping 算法构建二维栅格地图,并基于贝叶斯推理构建语义地图。随后,文献[16]提出了一种轻量级的卷积神经网络模型,并用该模型进行目标检测,获取环境物体的语义信息,同时使用激光 SLAM 算法获取机器人位姿并构建二维栅格地图,最后使用贝叶斯估计更新物体的占有概率,将物体信息与环境信息融合起来构建二维栅格的语义地图。文献[17]提出了 SuMa++ 算法,该算法通过 RangeNet++ 网络^[18]对激光雷达进行语义分割,滤除动态物体,并基于 SuMa^[19]构建高精度的语义地图。

收稿日期: 2021-12-13; 修回日期: 2022-01-09; 录用日期: 2022-01-19

基金项目: 国家重点研发计划(2019YFB1310000)、湖北省重点研发计划(2020BAB098)、国家自然科学基金(51874217)

通信作者: *liuqi_xl@163.com

以上算法都需要利用视觉传感器对环境中的物体进行识别,获取环境中物体的三维点云信息,但使用深度相机时存在两个问题:1)深度学习算法对物体进行检测时容易出现误识别;2)深度相机获取的深度值容易受到外界干扰,存在不可控误差。室内环境中存在丰富且固定的墙角信息,对墙角进行区分有助于移动机器人进行更高层次的定位和导航工作。本文针对以上问题提出了一种激光与视觉融合识别并构建语义栅格地图的算法,算法整体框架如图 1 所示。首先对激光雷达击中点进行直线拟合,提取环境中的凹凸墙角信息,同时利用视觉以及电子罗盘对墙角进行方向性区分,得到丰富的墙角类别信息;然后利用凹凸墙角信息对柜式物体进行误识别去除,并用激光雷达深度信息修正深度相机获取的深度值,得到更加精确的物体语义信息;最后同步机器人位姿,并结合贝叶斯估计更新墙角类别和物体的占有概率,增量式地构建语义地图。本文通过重定位实验验证了该语义地图可以改善定位性能。

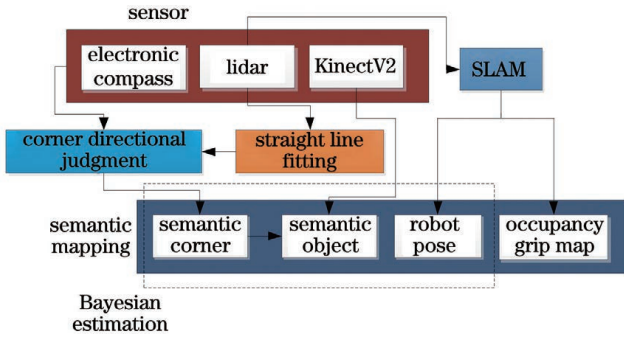


图 1 所提算法的整体框架图

Fig. 1 Overall framework of our algorithm

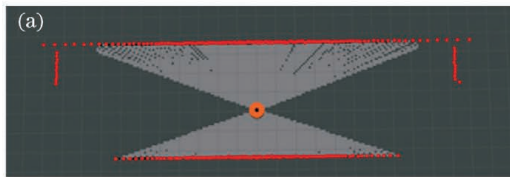
2 墙角类别区分

2.1 激光点直线拟合

激光雷达的一帧数据信息由一系列对障碍物的击中点组成,每个点的位置以极坐标形式进行保存。第 i 个激光点包含一个距离值 ρ_i 和角度值 θ_i ,其中 $\theta_i = f \cdot i$, f 为激光雷达的角度分辨率。将激光击中点的极坐标值转换到直角坐标系下,得到激光雷达的坐标 (x_i, y_i) 为

$$\begin{cases} x_i = \rho_i \cos \theta_i \\ y_i = \rho_i \sin \theta_i \end{cases} \quad (1)$$

在得到一帧激光数据之后,对每一帧激光点进行



直线拟合,拟合原理如图 2 所示,具体过程如下:

1) 对于一帧激光数据点 $Z(P_1, P_2, \dots, P_n)$, 根据端点坐标 P_1 和 P_n 求取直线 l , 计算中间激光点到直线 l 的最大距离 h_i 。

2) 若 h_i 大于阈值 H , 则在该点位置处将激光数据点拆分为 $Z_1(P_1, P_2, \dots, P_i)$ 和 $Z_2(P_i, P_{i+1}, \dots, P_n)$ 两段。

3) 对每一段激光数据 Z_i 重复上述步骤, 直至所有激光数据点的集合与对应直线的距离小于阈值 H , 此时将激光雷达数据分成了若干段。若激光数据的集合点数小于一定数量, 则将其去除。

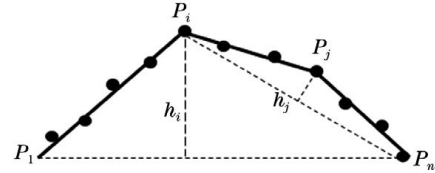


图 2 激光点直线拟合的原理

Fig. 2 Principle of laser point straight line fitting

对于获得的每个激光数据点的集合 $Z\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$, 利用线性最小二乘法进行直线拟合, 求取每段激光数据的直线方程, 其中误差方程可以表示为

$$e = (\mathbf{AX} - \mathbf{y})^T (\mathbf{AX} - \mathbf{y}), \quad (2)$$

式中: $\mathbf{A} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{bmatrix}$; \mathbf{y} 为一段激光数据点的值, $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$; \mathbf{X} 为直线方程的待求量, $\mathbf{X} = [a \quad b]$ 。

对式(2)所示的误差方程求偏导,并令其为 0, 则有

$$2\mathbf{A}^T \mathbf{AX} - 2\mathbf{A}^T \mathbf{y} = 0 \Rightarrow \mathbf{X} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}. \quad (3)$$

最后将每一段激光点的数据代入方程(3), 就可以求解线性方程的最优解 a 和 b 。

每一段激光点拟合的直线代表着激光击中面的直线信息, 但障碍物的长度是有限的, 因此只需要保存每段激光数据的两个端点值和对应的直线方程。图 3 所示为激光击中障碍物的数据点和进行直线拟合之后的激光线段。

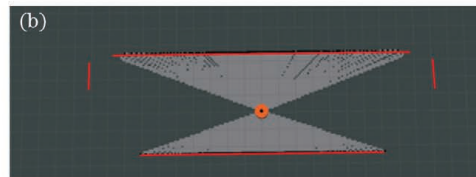


图 3 直线拟合。(a)激光击中点;(b)直线拟合

Fig. 3 Straight line fitting. (a) Laser hit point; (b) straight line fitting

2.2 激光直线对凹凸墙角的识别

一帧激光数据会提取到很多平面的直线方程,因此,只对相邻两条直线进行墙角判断。假设一条直线 l_1 的端点坐标分别为 $(x_{i_1}^1, y_{i_1}^1)$ 和 $(x_{i_1}^2, y_{i_1}^2)$, 另一条直线 l_2 的端点坐标为 $(x_{i_2}^1, y_{i_2}^1)$ 和 $(x_{i_2}^2, y_{i_2}^2)$, 若一对点之间的距离足够小, 小于设定的阈值 λ 时, 如式(4)所示, 就可以判断这两条激光拟合直线相邻。

$$\sqrt{(x_{i_1}^i - x_{i_2}^i)^2 + (y_{i_1}^i - y_{i_2}^i)^2} \leq \lambda, \quad (4)$$

式中: $i = j = 1, 2$; λ 可根据实际环境来定, 一般设置为 0.1 m。

接着, 判断两条相邻直线之间的区域是否存在墙角。设直线 l_1 的斜率为 a_1 , 直线 l_2 的斜率 a_2 , 如式(5)所示, 若两条相邻直线的斜率相乘的结果在 -1 左右, 则两直线的交点为墙角。

$$-1 - \beta < a_1 \cdot a_2 < -1 + \beta, \quad (5)$$

式中: β 为阈值, 一般设置为 0.1。

激光雷达识别到墙角之后, 对墙角按照凹凸类别进行区分, 一种是凸墙角, 另一种是凹墙角。如图 4(a) 所示, 蓝色圆点表示机器人位置, 激光雷达发射的红色激光束击中墙面, 得到距离信息。对凹凸墙角的区分可以根据激光击中点到机器人的距离变化进行判断。若两直线相交的墙角处距离机器人最近, 而越远离墙角位置的激光点到机器人的距离越远, 则判断为凸墙角; 若两直线相交的墙角距离机器人最近, 而越远离墙角位置的激光点到机器人的距离越近, 则判断为凹墙角。

图 4(b) 所示是识别出来的凹凸墙角。图中的红色线段是激光点拟合的直线, 蓝色圆点代表凸墙角, 绿色圆点代表凹墙角。

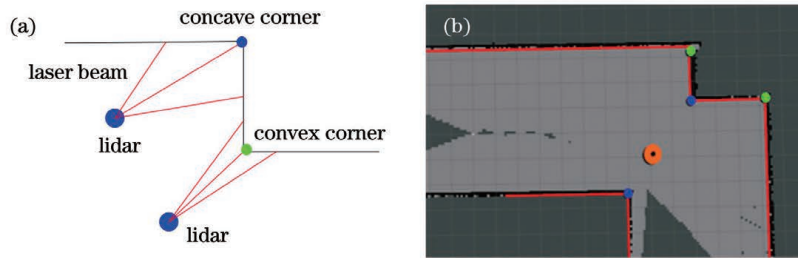


图 4 凹凸墙角识别。(a)凹凸墙角区分示意图;(b)凹凸墙角识别结果

Fig. 4 Recognition of convex and concave corners. (a) Schematic of distinguishing convex and concave corners; (b) recognition result of convex and concave corners

2.3 基于深度学习获取物体的语义信息

目标检测是对输入图片中的目标进行分类并确定目标位置, 一般分为基于候选框模式的目标检测和基于回归模式的目标检测。基于候选框模式的目标检测就是通常所说的“Two Stage”模式, 该检测方式的最大缺点就是运算速度慢。本文选用基于回归模式的目标检测算法 SSD^[20], 该算法结合了 YOLO^[21] 的网格化回归思想和 Faster R-CNN 中 RPN 网络的 Anchor 机制^[22], 并加入了多尺度特征图检测。SSD 算法能够同时兼顾检测速度和精度, 是目标检测算法中的较好选择。针对目标检测数据集, 本团队收集并制作了 8840 张包含墙角、门、柜子、椅子等 11 类室内场景下常见物体的图片, 其中的 7072 张用于训练, 1768 张用于测试。

语义分割是对输入图片中的每个像素点添加语义

标签, 最后得到一个与目标物体一样大的分类结果。基于目标检测的算法在获取物体检测框之后还要对物体背景进行去除, 一般采用 GrabCut 算法进行背景分割, 去除物体周围的背景信息, 获取目标物体的三维语义信息。基于语义分割的算法直接对物体进行像素级分割, 获取物体的三维语义信息。本团队选用 DeeplabV2 算法^[23] 对物体进行语义分割。DeeplabV2 在 DeeplabV1 的基础上将基础网络更换为 ResNet101^[24], 同时使用空洞空间卷积池化金字塔 (ASPP) 对输入使用不同的采样率进行空洞卷积采样, 既提高了特征图的分辨率, 又不会增加较大的计算量。针对语义分割数据集, 本团队收集制作了包含门、柜子、垃圾桶、椅子、床、背景共 6 个类别的 330 张图片, 其中 290 张用于训练, 40 张用于测试。语义分割算法和目标检测算法的结果如图 5 所示。



图 5 语义分割结果。(a)机器人视角;(b)语义分割;(c)目标检测

Fig. 5 Semantic segmentation results. (a) Robot perspective; (b) semantic segmentation; (c) target detection

对于室内环境下常见的椅子、门、柜子和垃圾桶等物体,使用 DeeplabV2 网络进行语义信息提取;由于墙角不便使用语义分割进行提取,因此采用 SSD 算法对墙角进行语义信息提取。

2.4 视觉对墙角进行方向性区分

激光对墙角的识别具有较高的准确性,但由于激光会同时识别到多个墙角,因此利用相机进行单一墙角的识别,并利用电子罗盘对墙角进行方向性区分。室内墙角具有 4 个朝向,每一类墙角都具有独特的标签,因此需要对墙角进行方向性判断,以获取墙角的类别标签。电子罗盘可以获得航向角,机器人搭载电子罗盘可以得到机器人的朝向,根据识别到的墙角对应的机器人朝向对墙角的方向性进行判断。本实验采用的电子罗盘型号为 DCM302B,在倾斜角小于 10° 时,其航向角误差为 $\pm 0.5^\circ$ 。针对电子罗盘易受到外界磁场干扰从而导致航向角精度不高的问题,本团队首先使用平面校准方法进行静态校准,然后利用椭圆假设法进行误差补偿,并采用最小二乘法求解补偿系数,完成对电子罗盘的误差补偿。

墙角方向性判断的示例模型如图 6 所示。正北方向的航向角为 0° ,航向角按照顺时针方向增大,角度范围在 $0^\circ \sim 360^\circ$ 之间。正北方向右侧的墙角设为墙角 1,按照逆时针顺序依次为墙角 2、墙角 3 和墙角 4。

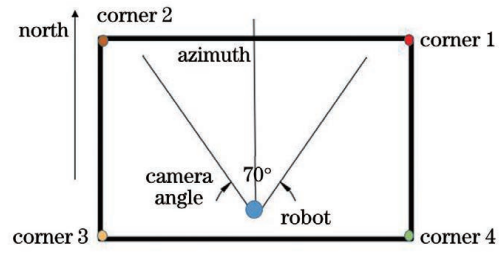


图 6 墙角类别和传感器示意图

Fig. 6 Schematic of wall corner categories and sensor

机器人搭载的相机为 KinectV2,其深度图最大视角范围是 70° ,机器人航向角位于相机视角范围的中间。

每类墙角都存在不重叠的航向角范围,在不重叠的航向角范围内识别到的墙角是对应的墙角类别。如图 7(a) 所示,在相机左极限视角范围内,相机只能识别到右上角的墙角 1,此时机器人航向角为 35° 。如图 7(b) 所示,在相机右极限视角范围内,相机也只能识别到右上角的墙角 1,此时的航向角为 55° 。因此,当机器人航向角在 $35^\circ \sim 55^\circ$ 之间时,相机视角只能识别到墙角 1。如图 7(c)、(d) 所示,机器人相机在左右极限视角下只能识别左上角的墙角 2,此时的航向角范围为 $305^\circ \sim 325^\circ$ 。根据类似的视角分析可以得到其他墙角对应的航向角,如表 1 所示。在对应的航向角范围内,相机识别到的墙角是对应类别的墙角。

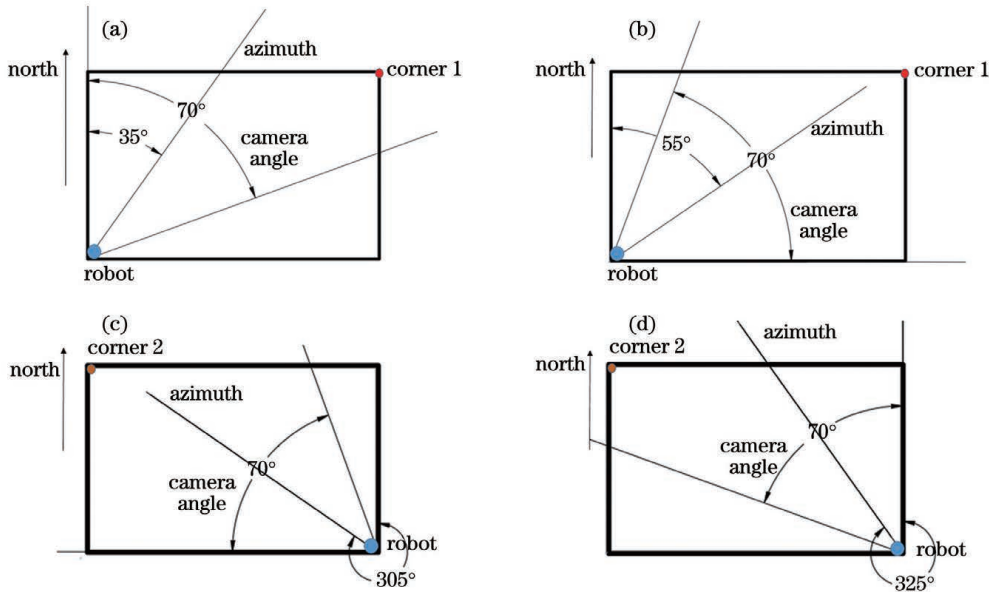


图 7 不重叠航向角下墙角类别的判断。(a)(c)左极限视角;(b)(d)右极限视角

Fig. 7 Wall corner category judgment under non-overlapping azimuth. (a)(c) Left limit view; (b)(d) right limit view

表 1 每类墙角对应的航向角

Table 1 Azimuth corresponding to each type of corner

Number of corner	Azimuth / ($^\circ$)
1	35-55
2	305-325
3	215-235
4	125-145

对于所有相邻的墙角,它们都具有重叠的航向角,如图 8 所示。对于墙角 1 和墙角 2,当航向角在 $325^\circ \sim 35^\circ$ 之间 ($325^\circ \sim 360^\circ$ 和 $360^\circ \sim 35^\circ$) 时,机器人相机能够识别到的墙角既可能是墙角 1 也可能是墙角 2。

对墙角 1 和墙角 2 在重叠航向角下进行类别判断的流程如下:

1) 首先判断墙角位于航向角的左侧还是右侧。若该侧只有一种墙角类型,则该墙角一定是对应的墙

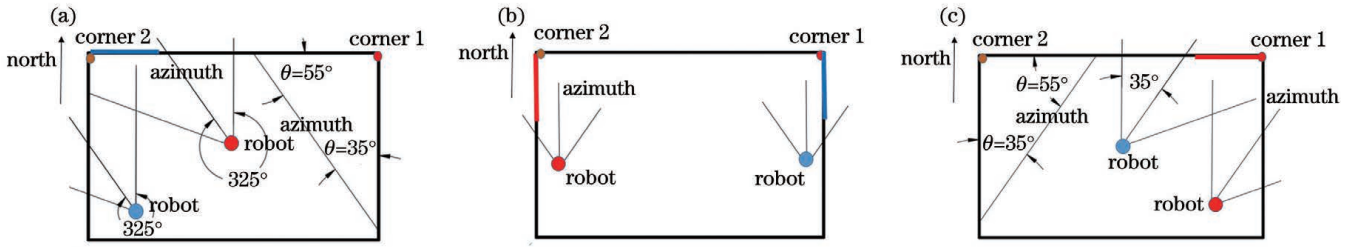


图 8 重叠航向角下的墙角类别判断。(a)航向角为 325°;(b)航向角为 0°;(c)航向角为 35°

Fig. 8 Wall corner category judgment under overlapping azimuth. (a) Azimuth is 325°; (b) azimuth is 0°; (c) azimuth is 35°

角类型。如图 8(a)、(b)所示,当航向角范围为 325°~360°时,墙角 1 只可能在航向角的右侧,墙角 2 可能在航向角的左侧也可能在右侧,因此出现在航向角左侧的墙角一定是墙角 2。如图 8(b)、(c)所示,当航向角范围在 0°~35°之间时,墙角 2 只可能在航向角的左侧,而墙角 1 可能在航向角的左侧也可能在右侧,因此出现在航向角右侧的墙角一定是墙角 1。

2) 如果该侧有两类墙角,则计算航向角角平分线与墙角的该侧墙面之间的夹角 θ 。对于图 8(a)、(b)中蓝色机器人位置,墙角 1 和墙角 2 都可能出现在航向角右侧,此时计算航向角角平分线与墙角右侧蓝色墙面之间的夹角 θ 。对于图 8(b)、(c)中红色机器人位置,墙角 1 和墙角 2 都可能出现在航向角左

侧,此时计算航向角角平分线与墙角左侧红色墙面之间的夹角 θ 。

3) 当 $\theta < 35^\circ$ 时,墙角类型为航向角一侧的墙角;当 $\theta > 55^\circ$ 时,墙角类型为航向角另一侧的墙角。对于流程 2) 中的图 8(a)、(b),墙角 1 和墙角 2 都可能出现在航向角右侧,若 $\theta < 35^\circ$,墙角类型为航向角右侧墙角 1;若 $\theta > 55^\circ$,墙角类型为航向角左侧墙角 2。对于流程 2) 中的图 8(b)、(c),若 $\theta < 35^\circ$,墙角类型为航向角左侧的墙角 2;若 $\theta > 55^\circ$,墙角类型为航向角右侧的墙角 1。

根据该判断方法可以得到重叠航向角下的墙角类型,如表 2 所示,用①、②、③、④表示对应的墙角 1~4,用左和右表示墙角位于航向角的左侧和右侧。

表 2 重叠航向角下的墙角类型

Table 2 Wall corner category under overlapping azimuth

θ	$\varphi = 325^\circ - 360^\circ$		$\varphi = 0^\circ - 35^\circ$		$\varphi = 235^\circ - 270^\circ$		$\varphi = 270^\circ - 305^\circ$	
	Left	Right	Left	Right	Left	Right	Left	Right
$\theta < 35^\circ$		①	②			②	③	
$\theta > 55^\circ$	②	②	①	①	③	③	②	②
θ	$\varphi = 145^\circ - 180^\circ$		$\varphi = 180^\circ - 215^\circ$		$\varphi = 55^\circ - 90^\circ$		$\varphi = 90^\circ - 125^\circ$	
	Left	Right	Left	Right	Left	Right	Left	Right
$\theta < 35^\circ$		③	④			④	①	
$\theta > 55^\circ$	④	④	③	③	①	①	④	④

Notes: φ represents azimuth.

为了验证在实际室内环境下对墙角的判断是否准确,在如图 9 所示的环境中进行墙角方向性判断,图中的方块砖是用来模拟墙壁的。机器人在图 9(a)中的位置时,识别到的是右边的墙角,如图 9(b)所示,此时结合电子罗盘数据和激光数据进行墙角类别的判断。图 9(c)是根据激光雷达数据拟合的墙面直线,图中机器人位于图片中心且方向朝北固定不动,因为墙角位于航向角的右侧,所以求取航向角角平分线与右侧墙面之间的夹角 θ ,根据墙角方向性判断方法得到该墙角类别为墙角 1,如图 9(d)所示。当机器人位于图 9(e)所示位置时,识别到的墙角如图 9(f)所示,此时墙角位于航向角的左侧,求取航向角角平分线与左侧墙壁之间的夹角 θ ,根据墙角方向性判断方法得到该墙角为墙角 2。

3 语义地图构建

3.1 激光雷达和相机联合标定

激光雷达和相机获取到的环境中障碍物的坐标都是相对于其自身坐标系下的坐标,如图 10 所示,点 P 在激光雷达坐标系下的坐标为 (X_l, Y_l, Z_l) ,在相机坐标系下的坐标为 (X_k, Y_k, Z_k) 。激光雷达坐标系与相机坐标系之间存在空间坐标变换关系,如式(6)所示,通过确定相机坐标系和激光雷达坐标系之间的关系,才能用激光雷达深度数据替代相机深度数据。

$$\begin{pmatrix} X_k \\ Y_k \\ Z_k \end{pmatrix} = \begin{pmatrix} \mathbf{R} & \mathbf{T} \\ \mathbf{0}^T & 1 \end{pmatrix} \begin{pmatrix} X_l \\ Y_l \\ Z_l \end{pmatrix}, \quad (6)$$

式中: \mathbf{R} 和 \mathbf{T} 分别为激光雷达坐标系与深度相机坐标

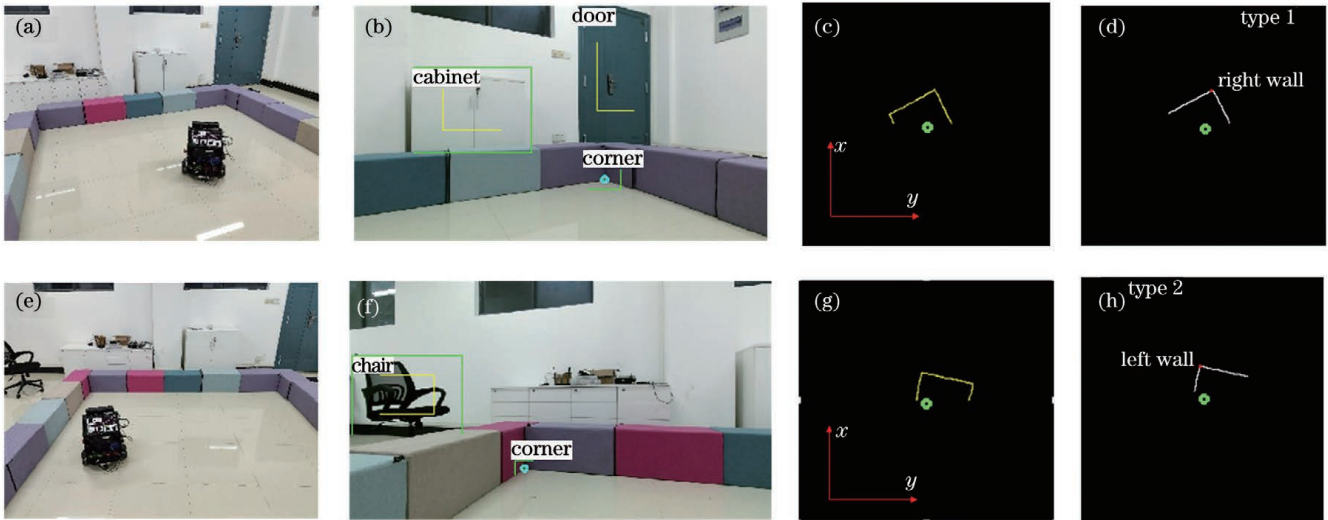


图 9 墙角方向性判断。(a)(e)机器人位置;(b)(f)机器人视角;(c)(g)激光拟合直线;(d)(h)墙角类别

Fig. 9 Wall corner directional judgment. (a)(e) Robot position; (b)(f) robot perspective; (c)(g) laser fitting line; (d)(h) wall corner category

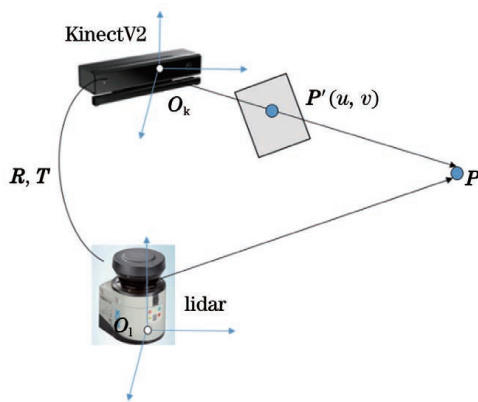


图 10 相机和激光雷达联合标定

Fig. 10 Joint calibration of camera and lidar

系之间的 3×3 旋转矩阵和 3×1 平移矩阵; $\mathbf{0}^T$ 为 $(0, 0, 0)$ 。

根据相机成像原理,空间坐标点在相机像素坐标系下成像,而相机的像素坐标点与相机坐标点之间的转换关系为

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \frac{1}{Z_k} \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_k \\ Y_k \\ Z_k \end{pmatrix}, \quad (7)$$

式中: $(u, v)^T$ 为像素坐标点; f_x, f_y, c_x, c_y 是相机内参,可通过相机标定得到; $Z_k = d/1000$, 其中 d 是像素点 $(u, v)^T$ 的深度值。

将式(6)代入式(7)可以得到激光雷达坐标点与像素坐标点之间的关系,如式(8)所示。在确定相机内参之后,将多组对应点代入式(8)可以求得激光雷达坐标系与相机坐标系之间的空间坐标变换关系,也就是旋转变换矩阵 \mathbf{R} 和平移变换矩阵 \mathbf{T} 。

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \frac{1}{Z_k} \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{R} & \mathbf{T} \\ \mathbf{0}^T & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \end{pmatrix}. \quad (8)$$

3.2 基于凸凹墙角修正柜式物体的语义

相机获取物体深度信息时存在两个问题:1)深度学习算法对物体的检测存在一定误差,在数据集不够丰富的情况下,深度学习算法很容易出现误识别;2)深度相机获取的深度值易受到外界干扰,存在一定的误差。针对相机存在的问题,本文利用激光雷达的精确性来修正相机误差。

本文使用的 KinectV2 相机的视角范围是 70° , 激光雷达扫描范围是 $0^\circ \sim 180^\circ$, 为了使激光雷达和相机的视角范围相同,取中间段 $65^\circ \sim 125^\circ$ 范围内的激光数据辅助相机进行识别。

室内环境中通常有很多柜式物体,比如柜子、床和沙发等。它们的样式是长方形的,存在多个角,通过激光雷达识别到的凸凹墙角来辅助相机对物体的识别,并修正物体的深度值。当相机误识别到柜子时,同时激光雷达识别到墙角信息,此时判断墙角位置与柜子位置的距离,如果距离较远,可以认定相机进行了误识别,去除该误识别,便可得到柜子的三维点云信息。当相机正确识别到柜子时,同时激光雷达识别到墙角信息,此时判断墙角位置与柜子位置的距离,如果距离较近,可以判断相机识别正确,同时利用激光雷达的深度信息修正相机识别到的柜子的深度信息。

机器人相机视角如图 11(a)所示,机器人语义分割结果如图 11(b)所示。由于存在误识别,机器人将后面的墙面以及旁边的障碍物也误识别成柜子,此时将柜子语义信息映射到世界坐标系下得到图 11(c)所示的柜子的二维栅格位置。从图中可以看出,由于误识别的存在,柜子的语义信息表达不准确。

如图 12 所示,机器人在识别到柜子时,对激光雷达信息进行直线拟合,提取凸凹墙角。图 12(a)是激光雷达检测出来的墙角。根据墙角与柜子的距离信息,将墙面误识别得到的柜子语义信息进行去除,同时



图 11 纯语义分割获取到的柜子语义。(a) 机器人视角;(b) 语义分割结果;(c) 点云坐标映射结果

Fig. 11 Cabinet semantics obtained by pure semantic segmentation. (a) Robot perspective; (b) semantic segmentation result; (c) point cloud coordinate mapping result

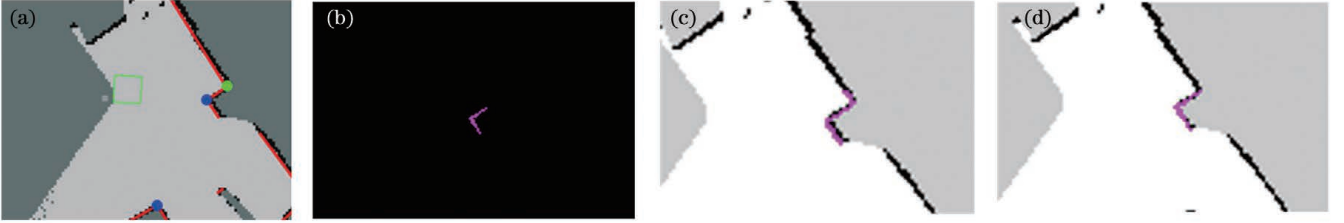


图 12 激光视觉融合得到的柜子语义。(a) 凸凹墙角的识别;(b) 修正之后的柜子语义;(c) 相机深度值;(d) 激光雷达深度值

Fig. 12 Cabinet semantics based on laser and vision. (a) Convex and concave wall corner recognition; (b) modified cabinet semantics; (c) camera depth value; (d) laser lidar depth value

使用激光雷达的深度信息替代相机识别到的柜子的深度信息,就可以实现凸凹墙角的识别。从图 12(b)中可以看出,使用激光雷达识别到的凸凹墙角来辅助相机对柜式物体进行识别,可以得到柜式物体正确的语义信息,柜子的语义信息表达准确。

通过柜子语义栅格与障碍物栅格地图的融合可以看出,使用深度相机得到的柜子深度值不精确,柜子信息(紫色)不能与黑色障碍物完全融合,如图 12(c)所示。使用激光雷达的深度信息替代相机识别到的柜子的深度信息,可以有效提高柜式物体的位置精度,图 12(d)中的柜子信息(紫色)可以与黑色障碍物完全融合,表示柜子的语义信息精确。

3.3 坐标转换以及贝叶斯估计构建语义地图

上述方法获取到的墙角和物体语义信息都位于激光雷达坐标系下,因此需要将激光雷达坐标系转化到世界坐标系下。在本次实验中,激光雷达位于机器人正前方,激光雷达坐标系和机器人坐标系之间只存在 x 轴方向上的平移,平移距离为 0.16 m。因此,激光雷达坐标系到机器人坐标系之间的转换矩阵 $T_{r,l}$ 以及机器人坐标系与世界坐标系之间的转换矩阵 $T_{w,r}$ 分别为

$$T_{r,l} = \begin{bmatrix} 1 & 0 & 0 & 0.16 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (9)$$

$$T_{w,r} = \begin{bmatrix} \cos \theta_t & -\sin \theta_t & 0 & x_t \\ \sin \theta_t & \cos \theta_t & 0 & y_t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (10)$$

式中: (x_t, y_t, θ_t) 是机器人在世界坐标系下的实时位

姿,可由 SLAM 算法得到。本文采用 Gmapping 算法进行机器人定位和环境栅格地图构建。Gmapping 算法是基于粒子滤波的 SLAM 算法,它将定位和建图分开,在进行机器人定位之后再行栅格地图构建;其定位过程就是采用一系列粒子的状态分布来代表机器人位姿的后验概率,然后通过求解粒子的位姿分布来确定机器人位姿 (x_t, y_t, θ_t) 。

将相机坐标系下的点坐标 (X_k, Y_k, Z_k) 转换到世界坐标系下的点坐标 (X_w, Y_w, Z_w) , 转换公式为

$$\begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = T_{w,r} T_{r,k} \begin{bmatrix} X_k \\ Y_k \\ Z_k \\ 1 \end{bmatrix}. \quad (11)$$

然后将点的坐标映射到二维栅格坐标系下(本文栅格地图的分辨率为 0.05 m, 计算结果向上取整), 计算公式为

$$\begin{pmatrix} X_g \\ Y_g \end{pmatrix} = \text{int} \left(\begin{matrix} X_w/0.05 \\ Y_w/0.05 \end{matrix} \right). \quad (12)$$

在构建物体语义栅格时,传感器存在误差和机器人存在打滑等因素使得连续帧获取的语义标签不一致,在建图时就表现为构建的语义地图重叠,地图中会重复累加物体的语义标签。机器人在位姿 1 得到的柜子和墙角语义如图 13 所示,当机器人移动到位姿 2 时,柜子和墙角的语义信息就会出现重叠。为了减小这些误差带来的语义重叠问题,本文使用贝叶斯估计从多个关键帧中获取语义标签,增量式地构建语义地图。给定机器人位姿为 $x_{1:t} = \{x_1, x_2, \dots, x_t\}$, 传感器数据为 $z_{1:t} = \{z_1, z_2, \dots, z_t\}$, 则语义地图 m 可表示为每一个语义栅格 m_i 的概率累积:

$$p(m | x_{1:t}, z_{1:t}) = \prod p(m_i | x_{1:t}, z_{1:t}). \quad (13)$$

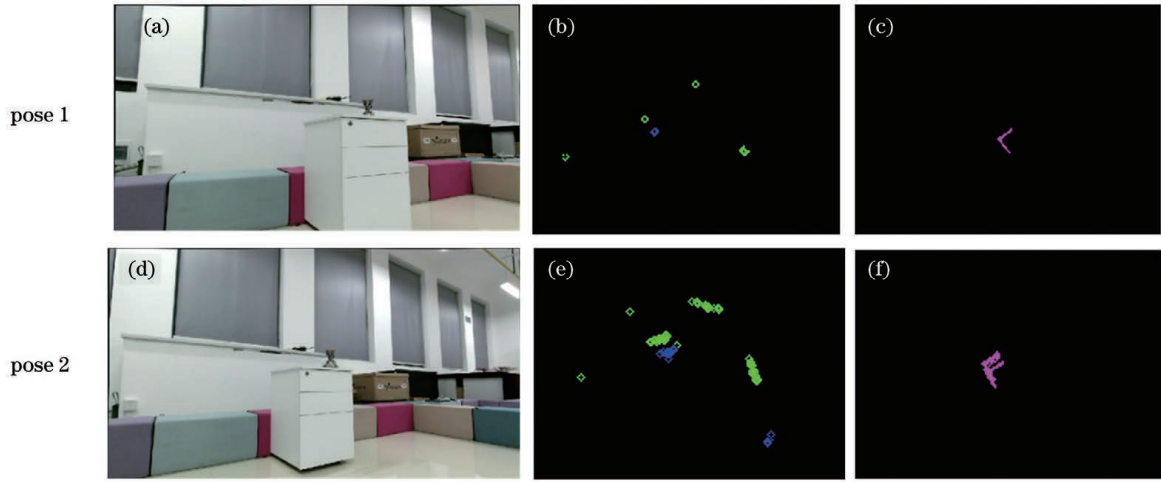


图 13 语义映射结果重叠。(a)(d) 机器人视角;(b)(e) 墙角语义映射;(c)(f) 物体语义映射

Fig. 13 Overlap of semantic mapping results. (a)(d) Robot perspective; (b)(e) wall corner semantic mapping; (c)(f) object semantic mapping

对每个语义栅格概率采用贝叶斯估计来更新语义栅格存在障碍物的概率,

$$p(m_i | x_{1:t}, z_{1:t}) = \frac{p(z_t | m_i, z_{1:t-1}, x_{1:t}) p(m_i | z_{1:t-1}, x_{1:t})}{p(z_t | z_{1:t-1}, x_{1:t})} = \frac{p(m_i | z_t, x_t) p(z_t | x_t)}{p(m_i)} \frac{p(m_i | z_{1:t-1}, x_{1:t-1})}{p(z_t | z_{1:t-1}, x_{1:t-1})} \quad (14)$$

同时可以得到非障碍物的概率为

$$p(\bar{m}_i | x_{1:t}, z_{1:t}) = \frac{p(\bar{m}_i | z_t, x_t) p(z_t | x_t)}{p(\bar{m}_i)} \frac{p(\bar{m}_i | z_{1:t-1}, x_{1:t-1})}{p(z_t | z_{1:t-1}, x_{1:t-1})} \quad (15)$$

将两者进行比较并取对数可以得到

$$\lg \frac{p(m_i | x_{1:t}, z_{1:t})}{p(\bar{m}_i | x_{1:t}, z_{1:t})} = \lg \left[\frac{p(\bar{m}_i)}{p(m_i)} \frac{p(m_i | z_{1:t-1}, x_{1:t-1})}{p(\bar{m}_i | z_{1:t-1}, x_{1:t-1})} \frac{p(m_i | z_t, x_t)}{p(\bar{m}_i | z_t, x_t)} \right] \quad (16)$$

式(16)可以简写成

$$l_{t,i} = l_{t-1,i} + l_{inv,i} - l_0, \quad (17)$$

式中: $l_{t,i}$ 表示当前时刻的语义栅格状态, $l_{t,i} = \lg \frac{p(m_i | x_{1:t}, z_{1:t})}{p(\bar{m}_i | x_{1:t}, z_{1:t})}$; $l_{t-1,i}$ 表示上一时刻的语义栅格状态, $l_{t-1,i} =$

$\lg \frac{p(m_i | z_{1:t-1}, x_{1:t-1})}{p(\bar{m}_i | z_{1:t-1}, x_{1:t-1})}$; $l_{inv,i}$ 表示传感器的逆观测模型, $l_{inv,i} = \lg \frac{p(m_i | z_t, x_t)}{p(\bar{m}_i | z_t, x_t)}$; l_0 表示先验概率, $l_0 =$

$\lg \frac{p(m_i)}{p(\bar{m}_i)}$ 。初始时刻语义栅格的占用概率和空闲概率都是 0.5,因此 $l_0 = 0$ 。

由式(17)可知,当前时刻语义栅格状态是上一时刻语义栅格状态和传感器观测概率的累加。机器人遍历整个室内环境,不断获取物体和墙角的语义信息,只有当语义栅格状态的最终后验概率大于设定阈值时才确定语义栅格存在,得到最终的语义标签,构建语义地图。

4 实验结果与分析

4.1 墙角类别的获取

在 Ubuntu16.04 系统下利用 ROS 操作系统,采用 Gazebo 和 Rviz 进行联合仿真,对墙角识别算法进行仿真实验验证。机器人在如图 14(a)所示的仿真环境下运动。首先利用 Gmapping 算法构建环境栅格地

图,然后通过激光直线拟合和凸凹墙角判断,获取凸凹墙角类别。图 14(b)~(d)是机器人移动过程中每个位姿下识别到的凸凹墙角,绿色代表凹墙角,蓝色代表凸墙角。通过遍历整个环境,得到图 15(a)、(b)所示的栅格地图和凸凹墙角类别。最后,将凸凹墙角坐标点进行坐标变换并与机器人位姿同步之后得到图 15(c)所示的带有凸凹墙角类别信息的栅格地图。从图中可以看出凸凹墙角识别准确,墙角位置也识别正确。

为了验证真实环境下所提算法构建语义地图的可行性,本课题组搭建了图 16(a)所示的移动机器人平台。该机器人主要搭载了微软的 KinectV2 深度相机和 Sick 公司的二维激光雷达 Lms111,其最大测量距

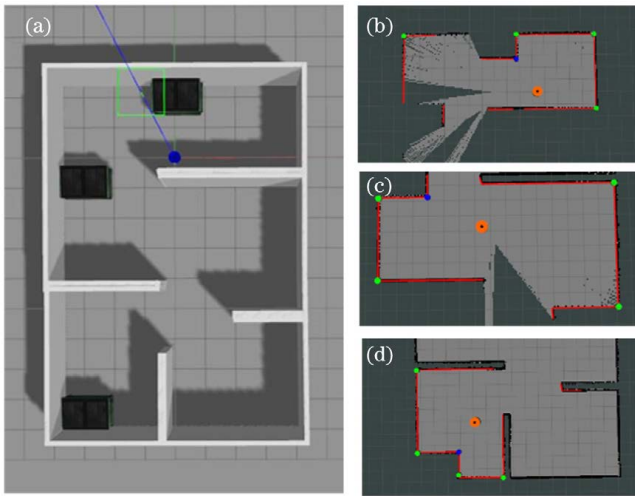


图 14 仿真环境下的凸凹墙角识别。(a)仿真环境；(b)~(d)凸凹墙角识别

Fig. 14 Convex and concave wall corner recognition in simulation environment. (a) Simulation environment; (b)~(d) convex and concave wall corner recognition

离为 20 m,角度分辨率为 0.25° ;机器人主控系统为酷

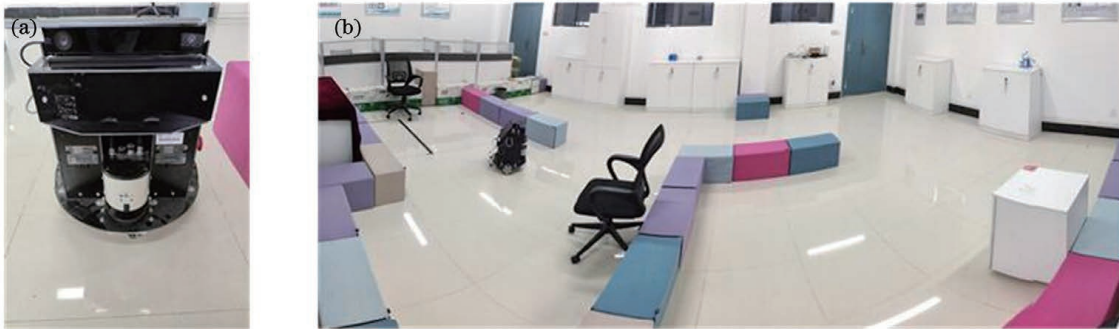


图 16 移动机器人平台和真实环境。(a)移动机器人平台;(b)真实环境

Fig. 16 Mobile robot platform and real environment. (a) Mobile robot platform; (b) real environment

表 3 墙角类别对应的颜色表

Table 3 Corresponding color table of corners

Wall corner	Convex wall corner	Concave wall corner 1	Concave wall corner 2	Concave wall corner 3	Concave wall corner 4
Color					

表 4 物体类别对应的颜色表

Table 4 Corresponding color table of objects

Object	Cabinet	Door	Chair	Trash can
Color				

首先控制机器人在室内环境中移动,使用 Gmapping 算法构建环境的栅格地图,同时利用激光雷达信息进行直线拟合,提取凸凹墙角的类别信息;然后使用 KinectV2 相机和电子罗盘进行凹墙角的方向性判断,得到 4 类不同的凹墙角类别信息;最后通过坐标变换以及贝叶斯估计增量式地构建墙角语义图。遍历整个环境得到了图 17(a)所示的栅格地图、图 17(b)所示的凸凹墙角类别和图 17(c)所示的 4 类凹墙角。将得到的凸墙角和 4 类凹墙角与栅格地图进行融合,

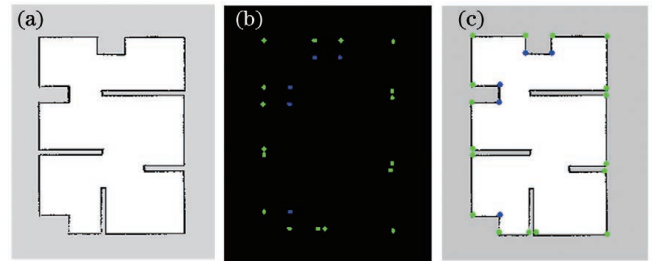


图 15 凸凹墙角识别。(a)栅格地图;(b)凸凹墙角类别;(c)融合墙角地图

Fig. 15 Recognition of convex and concave corners. (a) Grid map; (b) convex and concave wall corner categories; (c) fused wall corner map

睿 M4I7-D 迷你主机,并配有 ROS 机器人系统,同时配置有 Caffe 深度学习框架。

室内整体环境如图 16(b)所示。在环境中摆放方块砖,将环境分隔成若干个室内结构,同时在环境中摆放椅子、柜子等室内常见物体。对于每一类物体的语义信息,采用不同的颜色进行区分,得到表 3、4 所示的墙角和物体的语义信息表。

得到图 17(d)所示的带有凸墙角和 4 类凹墙角类别信息的语义栅格地图。

4.2 语义地图构建

机器人在构建物体语义地图时利用语义分割算法进行物体的识别,获取物体的语义信息。针对柜式物体,利用激光雷达识别出来的墙角信息对柜式物体的语义信息进行修正,获得精确的柜式物体的语义信息。

在图 16(b)所示的真实室内环境下,首先使机器人遍历环境,使用 Gmapping 算法构建环境的栅格地图,构建的栅格地图如图 17(a)所示;然后对激光雷达击中点进行直线拟合,提取环境中的凸凹墙角信息,同时利用深度相机提取环境中物体的语义信息,利用凸凹墙角信息对柜式物体进行误识别去除,并利用激光

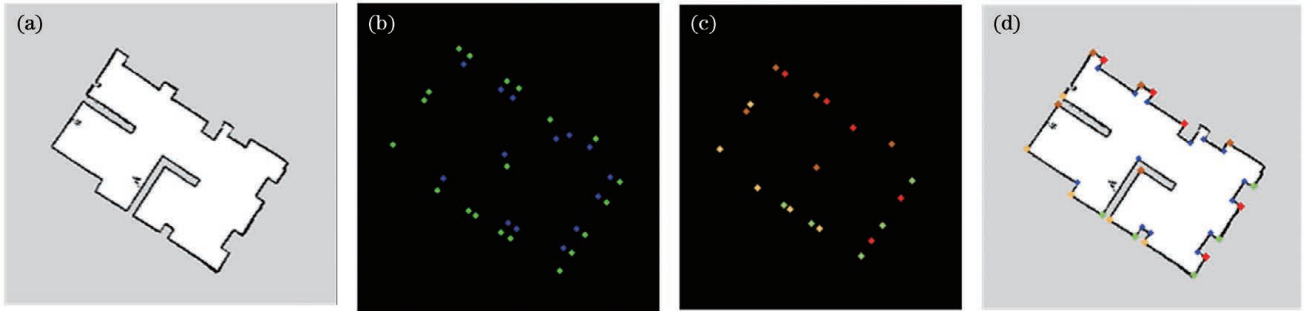


图 17 凸凹墙角的识别。(a)栅格地图;(b)凸凹墙角类别;(c)4 类凹墙角;(d)融合后的墙角地图

Fig. 17 Recognition of convex and concave corners. (a) Grid map; (b) convex and concave wall corner categories; (c) four types of concave corners; (d) fused wall corner map

雷达深度信息修正深度相机获取的深度值,得到更加精确的物体语义信息;再结合贝叶斯估计更新凸凹墙角和物体的占有概率,增量式地构建语义地图,得到如图 18(a)所示的物体语义地图;接着将物体的语义地图与栅格地图进行有效融合,得到带有物体类别信息的语义地图,如图 18(b)所示;最后将图 17(b)、(c)中的墙角类别信息与物体语义栅格地图进行融合,得

到最终的语义地图,如图 18(c)所示。从图 18(c)中可以看出物体轮廓基本与栅格障碍物区域吻合,不同的墙角类别也对应正确的墙角位置处。该语义地图正确反映了环境中的语义信息,有利于机器人正确地理解环境。

在图 19(a)所示的另一个真实环境下,机器人遍历整个环境构建了环境的栅格地图,如图 19(b)所示,

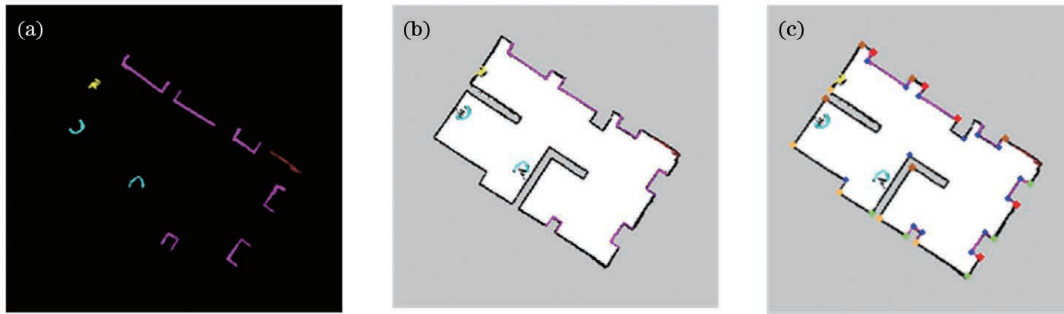


图 18 语义地图。(a)物体语义地图;(b)物体语义栅格地图;(c)最终的语义栅格地图

Fig. 18 Semantic map. (a) Object semantic map; (b) object semantic grid map; (c) final semantic grid map

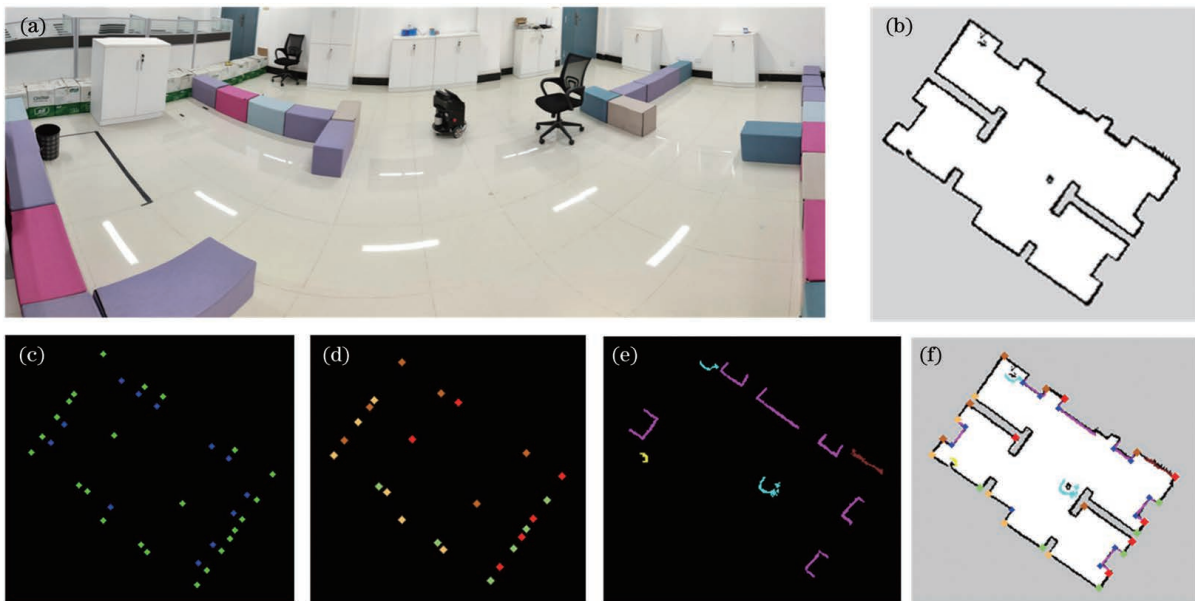


图 19 真实环境以及语义地图。(a)真实环境;(b)栅格地图;(c)凸凹墙角类别;(d)4 类凹墙角;(e)物体语义地图;(f)最终的语义栅格地图

Fig. 19 Real environment and semantic map. (a) Real environment; (b) grid map; (c) convex and concave wall corner categories; (d) four types of concave corners; (e) object semantic map; (f) final semantic grid map

环境地图与栅格地图一一对应;同时,利多传感器进行识别区分得到了图 19(c)所示的凸凹墙角类别和图 19(d)所示的 4 类凹墙角。利用深度相机提取环境中的物体语义信息,利用凸凹墙角信息对柜式物体进行误识别去除,并利用激光雷达深度信息修正深度相机获取的深度值,得到更加精确的物体语义信息;然后结合贝叶斯估计更新物体的占有概率,增量式地构建语义地图,得到的物体语义地图如图 19(e)所示;最后将墙角类别信息与物体语义地图进行融合,就可以得到最终的语义栅格地图,如图 19(f)所示。

4.3 利用语义地图改善相似场景下的定位性能

为了验证本文算法构建的语义地图的实用性,使用语义地图进行重定位实验,并与原始 AMCL 算法的定位结果进行对比。

移动机器人在相似环境中容易丢失自身位姿,重定位过程也难以成功。本团队在图 19(a)所示的相似环境中进行重定位对比实验,验证利用语义地图进行重定位的实用性和有效性。相似环境对应的栅格地图如图 20(a)所示,利用本文算法构建的语义地图如图 20(b)所示。从图中可以看到区域①和区域②的环境特征极其相似,机器人通过旋转运动能较快地获得周围信息,因此在区域②中进行旋转运动,完成重定位对比实验。

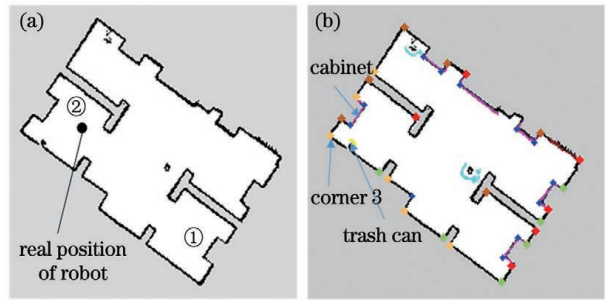


图 20 环境地图。(a) 栅格地图;(b) 语义地图
Fig. 20 Environment map. (a) Raster map;
(b) semantic map

移动机器人在上述地图和位置处进行重定位时,原始 AMCL 算法的粒子变化过程如图 21 所示。定位开始时,在地图中均布粒子,如图 21(a)所示,然后控制机器人旋转定位,机器人通过激光雷达观测模型与地图的匹配概率来更新粒子权重,重复进行重采样之后,粒子分布如图 21(b)、(c)所示。由于房间①和房间②的布局极其相似,利用激光雷达观测模型进行匹配并计算粒子权重时,粒子权重较为均衡,因此重采样之后不能迅速地将“假位姿”的粒子去除掉,最终粒子收敛到两个相似的房间内,如图 21(d)所示。经过多次实验,粒子也会收敛至如图 21(e)、(f)所示的错误位置。

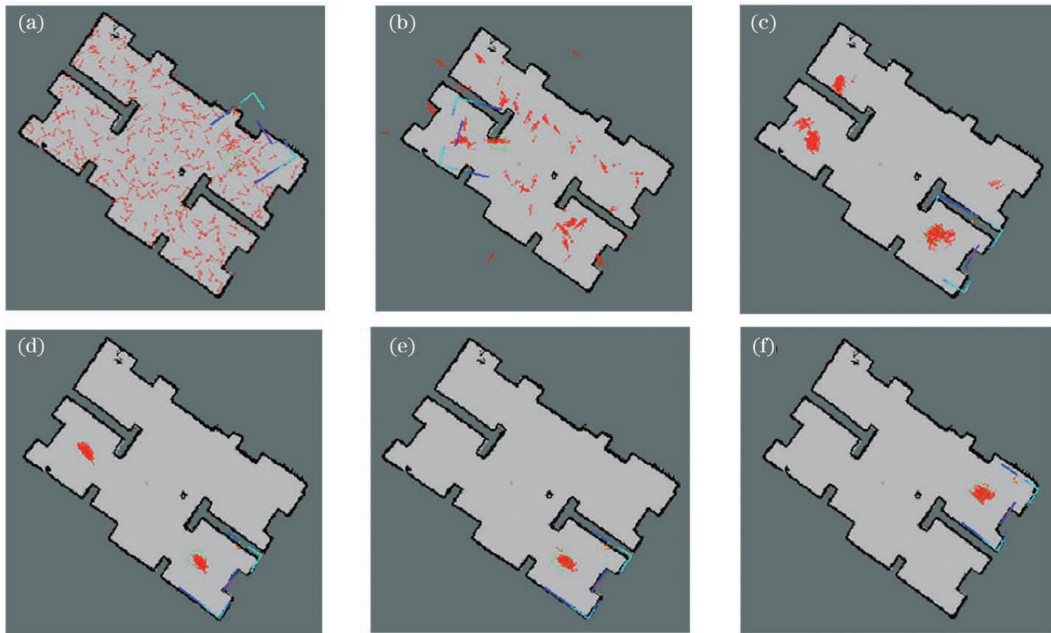


图 21 原始 AMCL 算法的粒子收敛过程

Fig. 21 Particle convergence process of original AMCL algorithm

利用语义地图进行重定位的过程如图 22 所示。开始定位时的粒子分布如图 22(a)所示,此时粒子的权重只根据激光雷达观测模型与地图匹配概率进行更新。当机器人识别到环境中的物体后,根据语义信息与语义地图的匹配概率来更新粒子的权重,重采样之后的粒子分布如图 22(b)所示。随着机器人不断运动,粒子快速收敛,最终粒子收敛至图 22(d)所示的状

态,机器人定位成功。

基于语义信息更新粒子权重的过程如图 23 所示。当 KinectV2 相机识别到环境中的语义信息时,利用语义地图进行机器人预定位,获取机器人的粗位姿。如图 23(a)所示,机器人识别到前方的“墙角”“柜子”“垃圾桶”,此时通过墙角方向性判断确定识别到的墙角为“墙角 3”,如图 23(b)所示。在识别到墙角之后,

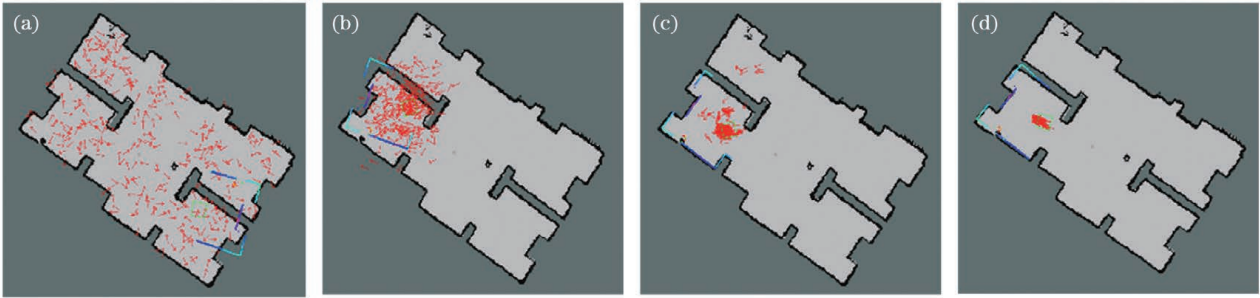


图 22 利用语义地图进行定位时的粒子收敛过程

Fig. 22 Particle convergence process in localization using semantic map

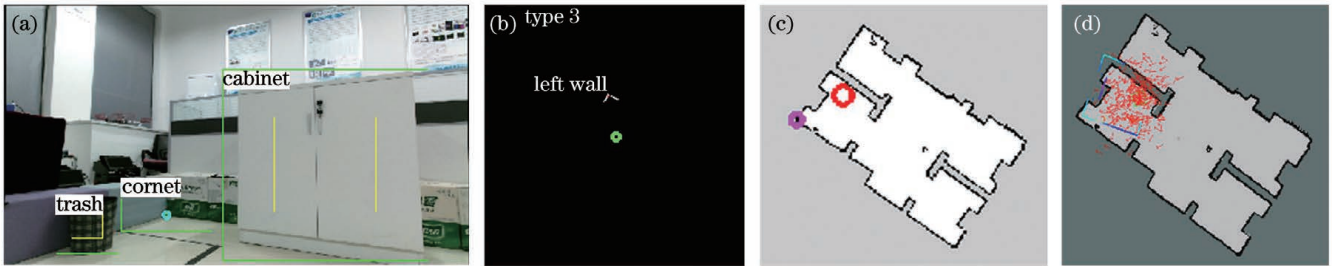


图 23 基于语义信息进行机器人预定位示意图。(a)机器人视角;(b)墙角类别;(c)墙角和机器人位置;(d)粒子分布

Fig. 23 Schematics of robot pre-location based on semantic information. (a) Robot perspective; (b) wall corner category; (c) position of wall corner and robot; (d) particle distribution

通过对照构建的语义地图,确定地图中周围存在“柜子”“垃圾桶”的“墙角 3”是图 23(c)所示的洋红色圆圈标识的墙角,也就是图 20(b)中标识的“墙角 3”。在确定识别到的墙角位置之后,利用识别到的物体相对于机器人的位姿来反推机器人在地图中的位姿,最终确定机器人的可能位姿如图 23(c)中的红色圆圈所示。在得到机器人粗位姿的先验值之后,将粒子的权重根据每一个粒子与机器人粗位姿的距离进行高斯展开,距离机器人粗位姿越近的粒子,其权重越大,经过重采样之后,权重小的粒子被去除,权重大的粒子进行复制,最后得到如图 23(d)所示的粒子分布。利用语义地图反推机器人位姿可以为机器人提供一个先验值,这样就可以避免相似环境下粒子权重较为均衡导致很难收敛正确的情况,从而提高了机器人重定位的鲁棒性。

为了进一步说明利用语义地图进行重定位可以有效提高重定位的鲁棒性,本团队在同等条件下进行了 20 组重定位实验,实验结果如表 5 所示。由于环境布局极其相似,原始 AMCL 利用激光雷达观测模型与栅格地图匹配概率来更新粒子权重,很难进行区域区分,粒子不能准确收敛,最后重定位的成功率为 25%。加载语义地图进行重定位可以提供一个先验位姿,粒子聚集在正确的房间下,机器人很快收敛成功,重定位的成功率为 95%。失败的案例是由于出现了误识别,机器人位姿估计不准确,因此定位失败。由以上数据可以看出利用语义信息进行机器人重定位可以有效提高重定位的鲁棒性。

表 5 不同方法的重定位实验结果对比

Table 5 Comparison of relocation experimental results of different methods

Location method	The number of successful location	Success rate / %
Location without semantic map	5	25
Location with semantic map	19	95

为了对比原始 AMCL 算法重定位的速度与加载语义地图进行重定位的速度,在两种方法都能够成功定位的情况下进行比较,并且记录每次重采样之后粒子的均权重。图 24 所示是两种方法的粒子均权重变换曲线。从加载语义地图的定位曲线可以看出,迭代到第 7 次后,粒子均权重迅速上升,最后在迭代 14 次之后居于稳定水平,此时粒子成功收敛。在原始 AMCL

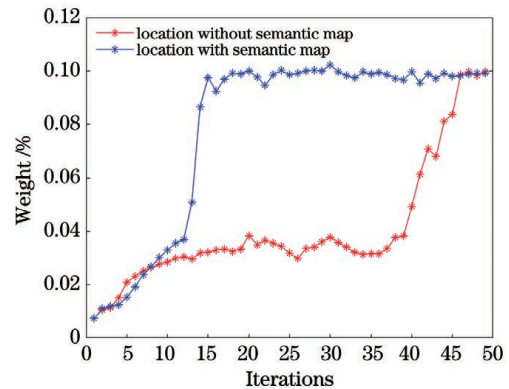


图 24 粒子收敛速率对比

Fig. 24 Particle convergence rate comparison

算法的重定位过程中,由于房间环境极其相似,粒子均权重一直不能上升,直至迭代第 37 次之后,粒子均权重才开始上升,直至迭代到第 45 次之后,粒子均权重居于稳定水平,粒子收敛成功。

该实验表明,在同等条件下,利用语义地图进行全局重定位可以有效提高定位速度。综上,利用语义地图可以有效提高定位的鲁棒性以及定位速度,说明该语义地图能有效改善定位性能。

5 结 论

为了提高语义地图的精度和丰富性,本团队利用激光雷达传感器进行激光点直线拟合,提取环境中的凸凹墙角信息,同时利用视觉及电子罗盘对凹墙角进行方向性区分,得到 4 个不同朝向的凹墙角;接着利用凸凹墙角信息进行相机误识别去除,并利用激光深度值信息修正相机的深度值,得到正确的柜式物体语义信息;然后将得到的语义信息进行坐标变换,并结合贝叶斯估计方法增量式地构架物体语义地图;最后将得到的语义地图和栅格地图进行融合得到了最终的语义栅格地图。本文通过仿真环境以及真实环境下的实验证明了该方法的有效性,同时利用该语义地图进行重定位实验,以验证该语义地图的实用性。语义分割的精度会直接影响语义地图构建的精度,如何提高语义分割算法的准确率是下一步的研究重点。在动态环境下,移动机器人容易受到运动物体的影响,导致语义信息提取不准确,因此,如何进行动态障碍物的去除也是下一步的研究重点。

参 考 文 献

- [1] Cadena C, Carlone L, Carrillo H, et al. Past, present, and future of simultaneous localization and mapping: toward the robust-perception age [J]. *IEEE Transactions on Robotics*, 2016, 32(6): 1309-1332.
- [2] 迟健男, 徐心和. 移动机器人即时定位与地图创建问题研究 [J]. *机器人*, 2004, 26(1): 92-96.
Chi J N, Xu X H. Research on simultaneous localization and mapping of mobile robot [J]. *Robot*, 2004, 26(1): 92-96.
- [3] 蒋林, 向超, 朱建阳, 等. 加载语义似然估计的粒子滤波重定位 [J]. *电子学报*, 2021, 49(2): 306-314.
Jiang L, Xiang C, Zhu J Y, et al. Particle filter relocation with semantic likelihood estimation [J]. *Acta Electronica Sinica*, 2021, 49(2): 306-314.
- [4] 蒋林, 聂文康, 朱建阳, 等. 基于具有墙角信息的语义地图改进 AMCL 重定位算法 [J/OL]. *机械工程学报*: 1-12 [2022-01-03]. <http://kns.cnki.net/kcms/detail/11.2187.TH.20211203.1419.010.html>.
Jiang L, Nie W K, Zhu J Y, et al. Improved AMCL relocation algorithm based on semantic map with corner information [J/OL]. *Journal of Mechanical Engineering*: 1-12 [2022-01-03]. <http://kns.cnki.net/kcms/detail/11.2187.TH.20211203.1419.010.html>.
- [5] 张磊, 徐孝彬, 曹晨飞, 等. 基于动态特征剔除的图像与点云融合的机器人位姿估计方法 [J]. *中国激光*, 2022, 49(6): 0610001.
Zhang L, Xu X B, Cao C F, et al. Robot pose estimation method based on image and point cloud fusion with dynamic feature elimination [J]. *Chinese Journal of Lasers*, 2022, 49(6):

- 0610001.
- [6] 蒋林, 李峻, 马先重, 等. 一种改进骨架提取的 Voronoi 路径规划 [J]. *机械工程学报*, 2020, 56(13): 138-148.
Jiang L, Li J, Ma X Z, et al. Voronoi path planning based on improved skeleton extraction [J]. *Journal of Mechanical Engineering*, 2020, 56(13): 138-148.
- [7] 白云汉. 基于 SLAM 算法和深度神经网络的语义地图构建研究 [J]. *计算机应用与软件*, 2018, 35(1): 183-190.
Bai Y H. Research on semantic mapping based on SLAM algorithm and deep neural network [J]. *Computer Applications and Software*, 2018, 35(1): 183-190.
- [8] Xiang Y, Fox D. DA-RNN: semantic mapping with data associated recurrent neural networks [EB/OL]. (2017-03-09) [2021-06-08]. <https://arxiv.org/abs/1703.03098>.
- [9] Newcombe R A, Izadi S, Hilliges O, et al. KinectFusion: real-time dense surface mapping and tracking [C] // 2011 10th IEEE International Symposium on Mixed and Augmented Reality, October 26-29, 2011, Basel, Switzerland. New York: IEEE Press, 2011: 127-136.
- [10] Whelan T, Leutenegger S, Salas Moreno R, et al. ElasticFusion: dense SLAM without a pose graph [C] // *Robotics: Science and Systems XI, Robotics: Science and Systems Foundation*, July 13-17, 2015, Rome, Italy. [S.l.: s.n.], 2015.
- [11] McCormac J, Handa A, Davison A, et al. SemanticFusion: dense 3D semantic mapping with convolutional neural networks [C] // 2017 IEEE International Conference on Robotics and Automation, May 29-June 3, 2017, Singapore. New York: IEEE Press, 2017: 4628-4635.
- [12] Li X P, Belaroussi R. Semi-dense 3D semantic mapping from monocular SLAM [EB/OL]. (2016-11-13) [2021-08-06]. <https://arxiv.org/abs/1611.04144>.
- [13] 赵亮, 胡杰, 刘汉, 等. 基于语义分割的深度学习激光点云三维目标检测 [J]. *中国激光*, 2021, 48(17): 1710004.
Zhao L, Hu J, Liu H, et al. Deep learning based on semantic segmentation for three-dimensional object detection from point clouds [J]. *Chinese Journal of Lasers*, 2021, 48(17): 1710004.
- [14] 邹斌, 林思阳, 尹智帅. 基于 YOLOv3 和视觉 SLAM 的语义地图构建 [J]. *激光与光电子学进展*, 2020, 57(20): 201012.
Zou B, Lin S Y, Yin Z S. Semantic mapping based on YOLOv3 and visual SLAM [J]. *Laser & Optoelectronics Progress*, 2020, 57(20): 201012.
- [15] Liu Z Y, von Wichert G. Extracting semantic indoor maps from occupancy grids [J]. *Robotics and Autonomous Systems*, 2014, 62(5): 663-674.
- [16] 李秀智, 李尚宇, 贾松敏, 等. 实时的移动机器人语义地图构建系统 [J]. *仪器仪表学报*, 2017, 38(11): 2769-2778.
Li X Z, Li S Y, Jia S M, et al. System of real time mobile robot semantic map building [J]. *Chinese Journal of Scientific Instrument*, 2017, 38(11): 2769-2778.
- [17] Chen X, Milioto A, Palazzolo E, et al. SuMa++: efficient LiDAR-based semantic SLAM [C] // 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), November 3-8, 2019, Macao, China. New York: IEEE Press, 2019: 4530-4537.
- [18] Milioto A, Vizzo I, Behley J, et al. RangeNet++: fast and accurate LiDAR semantic segmentation [C] // 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), November 3-8, 2019, Macao, China. New York: IEEE Press, 2019: 4213-4220.
- [19] Behley J, Stachniss C. Efficient surfel-based SLAM using 3D laser range data in urban environments [C] // *Robotics: Science and Systems XIV, Robotics: Science and Systems Foundation*, June 26-30, 2018, Pittsburgh, Pennsylvania, USA. [S.l.: s.n.], 2018.
- [20] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector [M] // Leibe B, Matas J, Sebe N, et al. *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9905: 21-37.

- [21] Redmon J, Farhadi A. YOLOv3: an incremental improvement [EB/OL]. (2018-04-08) [2021-05-06]. <https://arxiv.org/abs/1804.02767>.
- [22] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [23] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
- [24] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

Identifying and Constructing Semantic Maps Based on Laser and Vision Fusions for Improving Localization Performance

Jiang Lin^{1,2}, Liu Qi^{1*}, Lei Bin^{1,2}, Zuo Jianpeng¹, Zhao Hui^{2,3}

¹Key Laboratory of Metallurgical Equipment and Control Technology, Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081, Hubei, China;

²Institute of Robotics and Intelligent Systems, Wuhan University of Science and Technology, Wuhan 430081, Hubei, China;

³Hubei Key Laboratory of Mechanical Transmission and Manufacturing Engineering, Wuhan University of Science and Technology, Wuhan 430081, Hubei, China

Abstract

Objective Map construction is crucial to simultaneous localization and mapping (SLAM) technology and the foundation for robot positioning and navigation tasks. The environment map constructed using the traditional SLAM approach only provides the environment's topological and geometric information; it fails to provide the object's semantic information, thereby resulting in the poor understanding of the environment. When global positioning is performed in a complicated environment, obtaining faster positioning using only the environmental scale information in the map is a challenge. A semantic map can realize a more intelligent human-computer interaction function of the robot and accelerate the robot repositioning. In the indoor environment, there is rich and fixed wall corner information. The distinction of wall corners aids the mobile robot in performing higher-level positioning and navigation work. An approach of incrementally constructing a semantic map based on the fusion of laser and vision to determine the semantic information of wall corners and objects is proposed. The findings reveal that using a semantic map can enhance the positioning performance of robots.

Methods First, the laser lidar hit point was fitted linearly. Concave and convex wall corner information in the environment was extracted. The camera and electronic compass were employed to judge several wall corner categories to achieve rich wall corner category information. Then, the concave and convex wall corner information was used to determine and remove the cabinet objects, and the depth value achieved by the depth camera was corrected using the depth information of lidar, to obtain a more accurate semantic information of the surrounding object. Finally, the robot pose was synchronized, Bayesian estimation was employed to update the wall corner category and the object's occupancy probability, and a semantic map was incrementally constructed.

Results and Discussions In the indoor environment, the category information of concave and convex wall corners is extracted. Green represents the concave wall corner and blue represents the convex wall corner (Fig. 4). Then, wall corners are identified via target detection. The electronic compass and laser lidar are integrated to judge the direction of wall corners to obtain wall corner category information in four directions. The identification findings are accurate in the actual environment (Fig. 9). Errors are prone to occur when obtaining semantic information about cabinet objects via semantic segmentation. Therefore, concave and convex wall corner information is employed to remove the false detection of semantic segmentation. The camera depth is corrected using lidar depth information (Fig. 12). Finally, the constructed object semantic map, grid map, and wall corner information are combined to obtain the final semantic map (Figs. 18 and 19). The object's outline is consistent with the grid obstacle area, and different wall corner categories correlates with the correct wall corner position. The semantic map represents the semantic information in the environment. The success rate of repositioning using the proposed algorithm is 95% (Table 5), and the speed of repositioning is three times faster than that of the approach without a semantic map (Fig. 24).

Conclusions Laser lidar sensors were used to linearly fit the laser lidar hit point. Concave and convex wall corner information from the environment was extracted. The vision and electronic compass were employed to distinguish the

concave wall's directional corners, and four concave wall corners with various orientations were obtained. Furthermore, concave and convex wall corner information was employed to remove camera misidentification. First, the camera depth value was corrected using the laser depth value to achieve the correct semantic information of the cabinet object. Then, the semantic information was converted to coordinates and the object's semantic map was incrementally constructed using Bayesian estimation. Finally, the semantic and grid maps were combined to obtain a final semantic grid map. Experiments in the simulation and real environment can prove the proposed method's effectiveness. The semantic map is employed for relocation experiments to validate the semantic map's practicality.

Key words remote sensing; straight line fitting; wall corner category; Bayesian estimation; semantic map; relocation