

# 基于多智能体的图优化 SLAM 构建方法

宋燕燕<sup>1</sup>, 秦 军<sup>1,2</sup>, 邢艳芳<sup>1</sup>, 汪晨曦<sup>1</sup>, 周洪萍<sup>1</sup>

(1. 中国传媒大学南广学院, 江苏 南京 211172;

2. 南京邮电大学, 江苏 南京 210003)

**摘 要:**随着人工智能和增强现实技术在社会中的地位稳步上升,这些领域的核心关键技术正在逐步实现突破,对于三维环境的动态实时的理解是当前增强现实技术研究方面最活跃的问题之一。为实现多人同时应用视觉同时定位与地图构建(simultaneous localization and mapping, SLAM)系统,将图优化框架的 SLAM 与多智能体进行结合研究,提出基于多智能体的 SLAM 构建方法。首先简要介绍了视觉 SLAM 框架,系统地分析了在构建三维场景时,利用相机将信息进行整合和预处理,并估算相邻图像之间的运动以及检测信息来构建整体的框架。从视觉中提炼出最优化的 3D 模型以及各种参数来达到三维重建,通过相机的运动过程来确定视觉的深度特性以及加强沉浸式的观感体验,最终根据其本身的可视化追踪和环境理解,将非线性优化方案结合多智能体进行 SLAM 构建,实现移动平台真实浏览虚拟样板间间的可视化与交互。

**关键词:**多智能体;图优化;同时定位与地图构建;三维重建;可视化

中图分类号:TP391.41

文献标识码:A

文章编号:1673-629X(2020)11-0205-05

doi:10.3969/j.issn.1673-629X.2020.11.038

## An Optimizing SLAM Construction Method Based on Multi-agent

SONG Yan-yan<sup>1</sup>, QIN Jun<sup>1,2</sup>, XING Yan-fang<sup>1</sup>, WANG Chen-xi<sup>1</sup>, ZHOU Hong-ping<sup>1</sup>

(1. Communication University of China, Nanjing, Nanjing 211172, China;

2. Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

**Abstract:** With the steady rise of the status of artificial intelligence and augmented reality technology in the society, the key technologies in these fields are gradually breaking through. The dynamic real-time understanding of 3D environment is one of the most active problems in the research of augmented reality. To realize the multi-person simultaneous application of simultaneous localization and mapping system, SLAM of graph optimization framework and multi-agent is combined and a multi-agent based SLAM construction method is put forward. Firstly, the visual SLAM framework is briefly introduced, and the whole frame is constructed by using the camera to integrate and preprocess the information and estimate the motion between the adjacent images and detect the information. Optimized 3D models and various parameters are extracted from vision to achieve three-dimensional reconstruction. The depth of vision is determined by the camera's motion process and the immersion experience is enhanced. Finally, according to its own visual tracking and environmental understanding, a multi-agent SLAM construction method based on nonlinear optimization is proposed, and visualization and interaction between the real browsing virtual templates on the mobile platform is presented.

**Key word:** multi-agent; graph optimization; simultaneous localization and mapping; three-dimensional reconstruction; visualization

## 0 引言

增强现实(augmented reality, AR)是一种结合虚拟化技术观察世界的方式,它利用计算机技术,将虚拟的物体实时地叠加到一个真实画面或空间,形成具有实时交互的三维图像画面,给用户带来更真实的体验与感受<sup>[1]</sup>。增强现实由于其在真实环境和虚拟环境中起到了一个连续体的作用,为相关的应用提供了一种

新的人机交互方式,满足用户在现实世界中真实地感受虚拟空间中模拟的事物,增强使用的趣味性和互动性。近年来,随着同步定位与地图构建(SLAM)算法的不断更新,已经成为一个研究热点问题,定位(运动追踪)和建图(稀疏或者稠密地图)构成了SLAM的核心功能。因此,SLAM可以说是AR中最基础的模块,属于设备感知周边环境的范畴<sup>[2-4]</sup>。

收稿日期:2020-01-16

修回日期:2020-05-20

基金项目:江苏省高等学校自然科学研究面上项目(18KJB520053, 18KJD520006, 16KJD20005);江苏高校哲学社会科学研究基金项目(2018SJA0761);江苏省大学生创新创业训练计划项目(201913687003Y);院级项目(2019KYJK004, 2019KYJG002)

作者简介:宋燕燕(1978-),女,硕士,副教授,研究方向为数字媒体技术。

在目前的研究中,SLAM 一方面充分利用图像信息来提取图像特征,另一方面也通过融合深度相机、IMU 等传感器满足视觉 SLAM 在困难条件时,仍然可以稳健定位建图的需要,在机器人自动化、智能化区域发挥着很重要的作用。在针对多机器人基于深度学习的 SLAM 算法中,国外学者提出了基于稀疏编码的迁移学习 SLAM 算法,这样就可以有效减小训练模型的数据量,并对模型的通用化进行了完善,使人们可以理解不同类型的数据,进而更好地对底层进行规划。在目前的研究中,对于避障问题,学者提出将导航策略转变为一个深度的神经网络,将障碍物信息设为输入量,速度控制信号设为输出量,通过大量数据训练模型,从而完成避障。异构多机器人搭载多种传感器相对于单一种类机器人具有更大的优势,基于多传感器数据融合 SLAM 的研究在搭载多机器人时可以更加有效地对复杂环境进行全面的地图构建<sup>[5-6]</sup>。

## 1 基于图优化的 SLAM 框架

视觉 SLAM 就是一个带有特定相机的承载主体,它可以在没有任何的提前探测的情况下,对自己接下来的路径进行规划。前端主要涉及到计算机视觉领域的问题,而后端主要与优化算法有关。

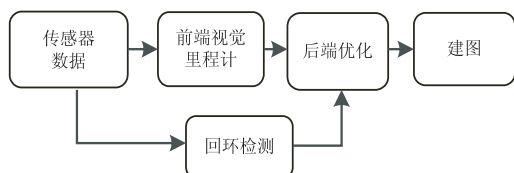


图 1 整体视觉 SLAM 流程

整个视觉 SLAM 框架如图 1 所示,主要包含五个步骤:传感器信息读取、视觉里程计、后端优化、回环检测、建图。图优化是把常规的优化问题采用图的方式来表示,图优化中的图就是数据结构里的图,一个图由若干个顶点以及连接着这些顶点的边组成,如图 2 所示,可以更直观地看到其结构。

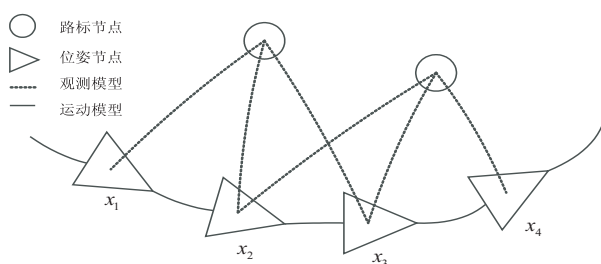


图 2 图优化例子

图优化 SLAM 方法是先通过前端的图构建得到初始状态估计值,在前端得到初始地图<sup>[7-8]</sup>。但由于传感器的噪声等,使得最后所得到的结果具有较大的误差,需要通过后端对初始状态进行估计来进一步优

化求解,使得地图的一致性和精度都得以提高。基于图优化的方法利用了这些信息来优化完整的地图路径,也称为全 SLAM 方法,如图 3 所示。

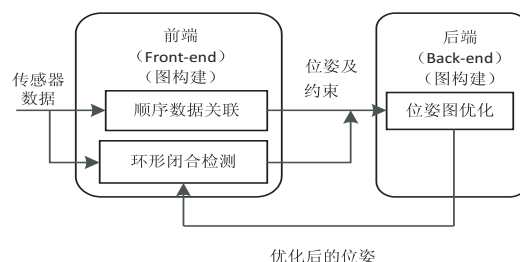


图 3 基于图优化的 SLAM 框架

## 2 SLAM 相机位姿的估计和优化

在 SLAM 中,除了对三维世界物体进行表述之外,在后端也要对在前端所得到的地图进行进一步的处理来提高它的精确度。需要描述当前的相机位姿,将它构建成一个优化问题,求解最佳值,使得误差最小化<sup>[9-10]</sup>。李代数对相机的位姿进行优化,两个李代数指数映射乘积的完整形式,由 BCH (Back-Campbell-Hausdorff) 公式给出。当处理两个矩阵指数之积时,会产生一些由二元运算组成的余项。在考虑到李代数  $\ln(\exp(\varphi_1^\wedge)\exp(\varphi_2^\wedge))^\vee$ ,当  $\varphi_1$  或  $\varphi_2$  为小量时,它们上面的项都可以忽略不计,因此可以得出 BCH 拥有线性近似表达:

$$\ln(\exp(\varphi_1^\wedge)\exp(\varphi_2^\wedge))^\vee \approx \begin{cases} J_1(\varphi_2)^{-1}\varphi_1 + \varphi_2 & \text{当 } \varphi_1 \text{ 为小量} \\ J_r(\varphi_1)^{-1}\varphi_2 + \varphi_1 & \text{当 } \varphi_2 \text{ 为小量} \end{cases} \quad (1)$$

当对一个旋转矩阵  $R_2$  (李代数为  $\varphi_2$ ) 左乘一个微小旋转矩阵  $R_1$  (李代数为  $\varphi_1$  时,可以近似地看作在原有的李代数上加上了一项  $J_1(\varphi_2)^{-1}\varphi_1$ 。在给近似添加了一个微小扰动的前提下,分成了左乘扰动和右乘扰动两种扰动方式。在使用时,须注意使用的是左乘模型还是右乘模型。

左乘 BCH 近似雅克比  $J_l$  为:

$$J_l = J = \frac{\sin\theta}{\theta}I + (1 - \frac{\sin\theta}{\theta})aa^T + \frac{1 - \cos\theta}{\theta}a^\wedge \quad (2)$$

它的逆为:

$$J_l^{-1} = \frac{\theta}{2}\cot\frac{\theta}{2}I + (1 - \frac{\theta}{2}\cot\frac{\theta}{2})aa^T - \frac{\theta}{2}a^\wedge \quad (3)$$

而右乘雅克比只需要将右乘中的自变量取负号即可得到:

$$J_r(\varphi) = J_l(-\varphi) \quad (4)$$

为了讨论位姿的导数,会引入构建位姿函数调整当前的值,以调整整体的估计值,使用李代数表示位姿,来对其进行求导,再对左乘或者右乘进行扰动,最后对其求导。这时把结果称为左扰动和右扰动模型。

### 3 非线性优化 SLAM 问题

即使相机的运动方程可以很好地符合针孔模型,但由于噪声的存在,数据依旧会受到很大的影响,运动方程和观测方程会对测量结果有偏差。所以需要在有噪声的数据中进行准确的状态估计,从噪声数据中恢复出所需要的东西,优化处理噪声数据<sup>[11]</sup>。

#### 3.1 状态估计问题

SLAM 模型一般由两个方程——运动方程和观测方程组成:

$$\begin{cases} x_k = f(x_{k-1}, u_k) + w_k \\ z_{k,j} = h(y_j, x_k) + v_{k,j} \end{cases} \quad (5)$$

观测方程在 SLAM 中具有特殊性,所以当在  $x_k$  对  $y_j$  进行观测,对应到图像中的像素位置  $z_{k,j}$ ,那么,便可得出观测方程:

$$sz_{k,j} = K \exp(\xi^\wedge) y_i \quad (6)$$

其中,  $s$  表示像素点之间的距离,  $k$  表示相机的内参。

在运动和观测设有的两个噪声项  $w_k$ ,  $v_{k,j}$  满足以下条件:

$$w_k \sim N(0, R_k), v_{k,j} \sim N(0, Q_{k,j}) \quad (7)$$

在噪声的干扰下,希望能够准确地计算出相机的位姿以及得到完整的地图(算出它们在有噪声的条件下的概率分布)。目前是采用非线性优化来解决这一状态估计问题,对每个时刻的数据中的噪点进行采集,然后对它进行状态估计。了解每个时刻中的误差点,这种方法已经成为了目前解决视觉 SLAM 中噪声问题的主要方法。在这里,把所有待估的变量放在一起:

$$x = \{x_1, \dots, x_n, y_1, \dots, y_n\}$$

所以可得到  $x$  的条件概率分布:  $P(x | z, u)$ , 在没有先验的情况下,没有告知刚体的运动方向,只有由像素点构成的图片时,就没有运动方程带来的具体数据,这样就只能依靠观测方程所产生的数据来判断,不必考虑在时间上的顺序。这时只能分析空间上物体的运动来得到三维结构信息,通过找到三维空物体图像的对点(在图像中找到角点特征),来恢复立体图像结构问题,这种方法被称为运动恢复结构 SFM(structure from motion)。为了进一步了解此时的条件分布特征,利用贝叶斯法则,可以得到:

$$P(x | z) = \frac{P(z | x)P(x)}{P(z)} \propto P(z | x)P(x) \quad (8)$$

其中,  $P(x)$  为先验。很难找到后验分布,因此需要先求解该状态下的最优估计。使得在该状态下后验概率最大化(maximize a posterior, MAP):

$$X_{\text{MAP}}^* = \operatorname{argmax} P(x | z) = \operatorname{argmax} P(z | x)P(x) \quad (9)$$

可以直接忽略贝叶斯法则的分母部分(与状态值  $x$  无关)。当无法判断刚体的位姿在什么地方时,就

明不能准确地得到先验概率。那么,此时可以求解  $x$  的最大似然估计:

$$X_{\text{MAP}}^* = \operatorname{argmax} P(z | x) \quad (10)$$

似然是指在当前的情况下可能得到的数据,所以最大似然估计的直接意义就是“在什么样的状态下,最有可能产生的现在观测到的数据”。

观测方程中,只有在  $x_i$  点可以看到  $y_i$  时,才会产生观测数据。由于视觉 SLAM 的特征点非常多,所以在实际中观测方程的量会远大于运动方程。如果认定没有运动方程,整个优化问题就只能由多个观测方程得到,相当于通过一组图像来恢复运动的结构。这就比较类似于 SFM,但 SLAM 中的图像有时间上的先后顺序,而 SFM 允许使用完全无关的图像。

#### 3.2 非线性优化方案

非线性优化最常见的两种方案是高斯牛顿法和列文伯格—马夸尔特方法。由于目标函数复杂,对问题提供不同的初始值都容易陷入局部最小值,所以使用高斯牛顿法和列文伯格—马夸尔特法优化初始值。

##### 3.2.1 高斯牛顿法

高斯牛顿法将  $f(x)$  进行一阶的泰勒展开:

$$f(x + \Delta x) \approx f(x) + J(x) \Delta x \quad (J(x) \text{ 为 } f(x) \text{ 的导数}) \quad (11)$$

它的目标是寻找下降矢量  $\Delta x$ , 使得  $\|f(x + \Delta x)\|^2$  达到最小,所以求解线性最小二乘:

$$\Delta x^* = \operatorname{argmin}_{\Delta x} \frac{1}{2} \|f(x) + J(x) \Delta x\|^2 \quad (12)$$

将目标函数对  $\Delta x$  求导,并令导数为零,可得:

$$J(x)^T J(x) \Delta x = -J(x)^T f(x) \quad (13)$$

求解的变量为  $\Delta x$ , 一般称为增量方程或者高斯牛顿方程(Gauss Newton equation)。将左边的系数项设为  $H$ , 右边定义为  $g$ :

$$H \Delta x = g \quad (14)$$

将  $J^T J$  作为二阶 Hessian 矩阵的近似,省略掉了计算  $H$  的过程。所以可得高斯牛顿法的步骤为:

(1) 给定初始值  $x_0$ 。

(2) 对第  $k$  次迭代,求出当前的雅克比矩阵  $J(x_k)$  和误差  $f(x_k)$ 。

(3) 求解增量方程  $H \Delta x = g$ 。

(4) 直到  $\Delta x_k$  足够小,则停止。否则,令  $x_{k+1} = x_k + \Delta x_k$ , 继续第二步。

从算法中可以看出,增量方程求解占据主要位置。它要求所用的近似  $H$  矩阵是可逆的(而且是正定的)。但是可能出现的  $J^T J$  却只有半正定性。在使用高斯牛顿法时,可能会出现  $J^T J$  为奇异矩阵或者病态的情况,导致算法不收敛。即使排除这一情况,在求步长  $\Delta x$  时,若它过大,也会使得局部近似不够准确,无法保证



它的迭性收敛。

### 3.2.2 列文伯格-马夸尔特法

高斯牛顿法在运算时具有局限性,它只能展开附近较好的近似效果,因此需要给  $\Delta x$  添加信赖区域(trust region)来提高它的准确性。在非线性优化中,这类方法被称为信赖区域法(trust region method)。一般区域的大小与近似模型和实际函数的差异成反比,用  $\rho$  来代替两个函数的差异性:

$$\rho = \frac{f(x + \Delta x) - f(x)}{J(x)} \quad (15)$$

其中,分子表示实际函数的下降值,分母表示近似模型的下降值。若  $\rho$  接近于 1,则近似值是正好的;若  $\rho$  太小,则近似性比较差,需要缩小范围;反之,则说明实际下降范围比较大,这时,需要放大范围。

因此构建了一个改良的非线性优化框架,即:

(1) 给定初始值  $x_0$  以及优化半径  $\mu$ 。

(2) 对于第  $k$  次迭代,可得:

$$\min_{\Delta x_k} \frac{1}{2} \|f(x_k) + J(x_k) \Delta x_k\|^2, \text{ s.t. } \|D \Delta x_k\|^2 \leq \mu$$

$\mu$  (这里的  $\mu$  为信赖半径)

(3) 计算  $\rho$ 。

(4) 若  $\rho > \frac{3}{4}$ , 则  $\mu = 2\mu$ 。

(5) 若  $\rho < \frac{1}{4}$ , 则  $\mu = 0.5\mu$ 。

(6)  $\rho$  大于阈值,则认为近似可行,令  $x_{k+1} = x_k + \Delta x_k$ 。

(7) 判断算法是否收敛,若不收敛,则继续第二步;否则结束。

马夸尔特提出将  $D$  取成非负数对角阵—通常用  $J^T J$  的对角元素平方根,这样使得在梯度维度上的约束范围会变得大一些。用拉格朗日乘子将它转化为一个无约束化问题:

$$\min_{\Delta x_k} \frac{1}{2} \|f(x_k) + J(x_k) \Delta x_k\|^2 + \frac{\lambda}{2} \|D \Delta x_k\|^2 \quad (16)$$

可以发现计算增量的线性方程仍然是该问题的核心:

$$(H + \lambda D^T D) \Delta x = g \quad (\text{以下用 } I \text{ 代替 } D^T D) \quad (17)$$

在这个增量方程中,相比于高斯牛顿法,多了一项  $\lambda I$ 。当  $\lambda$  较小时,  $H$  占主要地位;另一个方面,当  $\lambda$  较大时,  $\lambda I$  将占据主要地位。列文伯格—马夸尔特法为优化问题提供更稳定、准确的增量  $\Delta x$ ,在一定程度上避免了方程组的非奇异性 and 病态问题。

## 4 多智能体与 SLAM 构建

多智能体技术可以为 SLAM 提供很好的稳定性

和鲁棒性,使视觉 SLAM 得到更加有效可靠的特征。具有无监督学习机制的 SLAM 已经在现实应用中有很大的市场,将以上非线性优化方案结合多智能体技术使得实时追踪能够具有自身的协调能力,更好地应用于适应未来的市场发展机制<sup>[12-13]</sup>。视觉 SLAM 则是增强现实(AR)中一个很好的搭载辅助工具。它可以自主地识别空间中的墙壁,障碍物以及其他物品。当出现大规模的数据集时,基于数据的图优化 SLAM 能够自主提取特征,而非手动提取。但当出现大规模数据时,它的鲁棒性还需要进一步提升<sup>[14]</sup>。

基于深度学习图优化的 SLAM also 具有很强的灵活性,它不仅可以估计位置和构建三维点云地图,并且可以产生估计的不确定性,为后续的决策所服务<sup>[15]</sup>。但是它也存在误差很大,数据增长很快的情况。该文提供了一种解决方法:核函数。它能保证每条边的误差不会大到掩盖其他的边。把原先误差的二范数度量替换成一个增长没有那么快的函数,保证自己的光滑性质(否则无法求导),使得整个优化过程更为稳健,所以又叫鲁棒核函数(robust kernel)。

鲁棒核函数最常用的 Huber 核为:

$$H(e) = \begin{cases} \frac{1}{2} e^2 & \text{当 } |e| \leq \delta \\ \delta(|e| - \frac{1}{2}\delta) & \text{其他} \end{cases} \quad (18)$$

当误差  $e$  大于某个阈值  $\delta$  时,函数增加由二次形式变成了一次形式,相当于限制了梯度的最大值。如图 4 所示,可见误差较大时 Huber 核函数增长明显低于二次函数。在生成场景语义图时,如何进一步提高改善定位性和鲁棒性还需不断探索,期待场景语义图可以帮助定位,并进一步减少数据存储量,从而有助于大规模大尺度环境构图。

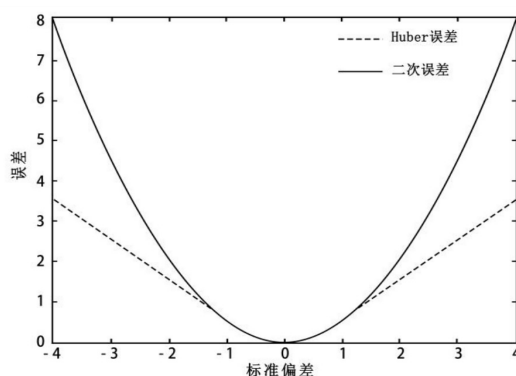


图4 Huber 核函数

通过该文构建的方法,将非线性优化方案结合多智能体进行 SLAM 构建,实现移动平台真实浏览虚拟样板间的可视化与交互,运动跟踪是通过识别和跟踪用户周围的可识别特征来完成的,然后利用这些点以及设备的朝向和加速度传感器来保持跟踪更新。如果

没有这些信息,就无法实现准确跟踪。随后通过 USB 或远程将设备与开发机器相连,然后在设备上构建和运行应用程序。绘制点云数据需要花费一段时间。并且随着点的个数的增加,在一个方向保持设备的时间就越长。这些点表示用于跟踪和解释环境的可识别特征点。放置对象用光线投射技术,光线投射技术可在两个维度上获取触摸点,并向场景投射光线,然后该射线测试场景中产生碰撞的其他对象。检测到平面的效果如图 5 所示,图 6 为点击平面后叠加出现另一个虚拟样板间的效果。

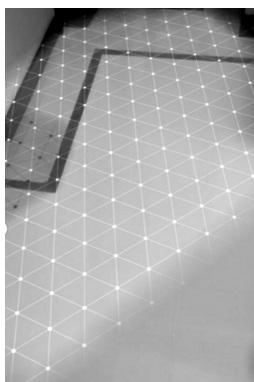


图 5 检测到平面



图 6 点击平面并出现门

## 5 结束语

基于图优化 SLAM 算法与多智能体结合,提出一种构建方法,并对增强现实进行应用讨论。现在人工智能和增强现实技术在社会中的地位稳步上升,这些领域的核心关键技术正在逐步实现突破,视觉识别技术也遥遥领先。图优化 SLAM 与多智能体的结合,给增强现实在构建完整地图提供了良好的开端,实现了 SLAM 与增强现实的结合。视觉 SLAM 系统在大尺度地图中的速度,优化算法能得到提升。在大环境下系统性将会变得脆弱,鲁棒性也随之降低。如果遇到特征点很少的场景、交叉的白平衡、图片失真等情况,这些给对特征检测增加很大的难题,数据会变大,导致无法很好地存储与处理地图。当将多智能体技术与其

相结合,通过智能体之间的相互协调性提高算法的速度,通过智能体之间的相互通信实现对每一个零散的点进行整合,能得到一个更好的三维地图,在未来的自动驾驶等领域会得到良好的应用。

## 参考文献:

- [1] 宋燕燕,曹效业,周 灵. 移动增强现实技术在互动展示中的应用研究[J]. 计算机技术与发展,2016,26(9):83-86.
- [2] DAVISON A J, REID I D, MOLTON N D, et al. Mono-SLAM: real-time single camera SLAM[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6):1052-1067.
- [3] 李 策,魏豪左,卢 冰,等. 基于深度视觉的 SLAM 算法研究与实现[J]. 计算机工程与设计,2017,38(4):1062-1066.
- [4] 揭云飞,王 峰,智凯旋,等. 视觉 SLAM 系统分析[J]. 电脑知识与技术,2018,14(19):221-223.
- [5] 朱笑笑,曹其新,杨 扬,等. 基于 RGB-D 传感器的 3D 室内环境地图实时创建[J]. 计算机工程与设计,2014,35(1):203-207.
- [6] 张文旭,马 磊,王晓东. 基于事件驱动的多智能体强化学习研究[J]. 智能系统学报,2017,12(1):82-87.
- [7] 王忠立,赵 杰,蔡鹤皋. 大规模环境下基于图优化 SLAM 的后端优化方法[J]. 哈尔滨工业大学学报,2015,47(7):20-25.
- [8] 卫 恒,吕 强,林辉灿,等. 多机器人 SLAM 后端优化算法综述[J]. 系统工程与电子技术,2017,39(11):2553-2565.
- [9] 宋鑫坤,陈万米,徐昱琳,等. 非结构化场景下移动机器人 FastSLAM 应用研究[J]. 计算机技术与发展,2010,20(2):95-98.
- [10] 王爱丽,胡长雨,韩 闯. 基于 ORB 特征的复杂场景下的增强现实[J]. 哈尔滨理工大学学报,2018,23(2):23-28.
- [11] MURARTAL R, MONTIEL J M M, TARDÓ S J D. ORB-SLAM: a versatile and accurate monocular SLAM system[J]. IEEE Transactions on Robotics, 2017, 31(5):1147-1163.
- [12] MOHAMED N, YUSOFF F, ISMAIL M, et al. The potential use of augmented reality in gamification[C]//International conference on computing & informatics. [s. l.]: [s. n.], 2015:159-167.
- [13] MURARTAL R, TARDOS J D. ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-D cameras[J]. IEEE Transactions on Robotics, 2017, 33(5):1255-1262.
- [14] 孙凤池,黄亚楼,康叶伟. 基于视觉的机器人同时定位与建图研究进展[J]. 控制理论与应用,2010,27(4):488-494.
- [15] 高 翔,张 涛,刘 毅,等. 视觉 SLAM 十四讲从理论到实践[M]. 北京:电子工业出版社,2017.