



(12) 发明专利申请

(10) 申请公布号 CN 111797688 A

(43) 申请公布日 2020.10.20

(21) 申请号 202010488128.2

G01C 11/02 (2006.01)

(22) 申请日 2020.06.02

(71) 申请人 武汉大学

地址 430072 湖北省武汉市武昌区珞珈山
武汉大学

(72) 发明人 姚剑 卓胜德 程军豪 龚烨
涂静敏

(74) 专利代理机构 武汉科皓知识产权代理事务
所(特殊普通合伙) 42222

代理人 胡琦旖

(51) Int. Cl.

G06K 9/00 (2006.01)

G06K 9/34 (2006.01)

G06K 9/62 (2006.01)

G06N 3/04 (2006.01)

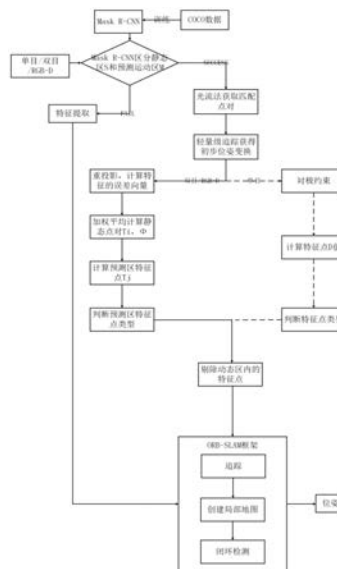
权利要求书3页 说明书8页 附图1页

(54) 发明名称

一种基于光流和语义分割的视觉SLAM方法

(57) 摘要

本发明属于视觉空间定位技术领域,公开了一种基于光流和语义分割的视觉SLAM方法,包括:采用语义分割网络对输入的图像信息进行分割,得到静态区域、预测动态区域;采用稀疏光流法对静态区域、预测动态区域进行特征追踪;对输入的图像信息中的特征点的类型进行判定,并移除动态特征点;将剔除运动特征点的集合作为追踪数据,输入至ORB-SLAM中进行处理,输出位姿结果。本发明解决了动态环境中SLAM追踪定位效果较差的问题,可以得到动态环境中位姿精度较高的轨迹信息。



1. 一种基于光流和语义分割的视觉SLAM方法,其特征在于,包括以下步骤:

步骤1、采用语义分割网络对输入的图像信息进行分割,得到静态区域、预测动态区域;

步骤2、采用稀疏光流法对所述静态区域、所述预测动态区域进行特征追踪;

步骤3、对输入的图像信息中的特征点的类型进行判定,并移除动态特征点;

步骤4、将剔除运动特征点的集合作为追踪数据,输入至ORB-SLAM中进行处理,输出位姿结果。

2. 根据权利要求1所述的基于光流和语义分割的视觉SLAM方法,其特征在于,所述步骤1中的所述输入的图像信息为单目相机对应的输入数据、双目相机对应的输入数据、深度相机对应的输入数据中的一种;

所述步骤3中,通过对极约束对单目相机对应的输入数据中的特征点的类型进行判断;通过重投影误差对双目相机对应的输入数据或深度相机对应的输入数据中的特征点的类型进行判断。

3. 根据权利要求1所述的基于光流和语义分割的视觉SLAM方法,其特征在于,所述步骤1包括以下子步骤:

步骤1.1、选取数据集对Mask R-CNN网络进行训练,得到训练好的语义分割网络;所述数据集包括多种类型数据作为潜在的运动对象;

步骤1.2、将图像信息输入至训练好的语义分割网络,完成图像分割,得到静态区域 A_s 、预测运动区域 A_m 。

4. 根据权利要求3所述的基于光流和语义分割的视觉SLAM方法,其特征在于,所述步骤2包括以下子步骤:

步骤2.1、利用稀疏光流法对所述静态区域 A_s 和所述预测运动区域 A_m 进行特征提取、匹配,得到静态匹配点对集合和预测运动匹配点对集合;

步骤2.2、基于SLAM的运行模型求解位姿。

5. 根据权利要求4所述的基于光流和语义分割的视觉SLAM方法,其特征在于,所述步骤2.2中的所述基于SLAM的运行模型求解位姿包括:

在 k 时刻,利用第 j 个路标 y_j 投影到当前帧得到投影位置 $h(\xi_k, y_j)$,得到对应的观测模型:

$$z_{k,j} = h(\xi_k, y_j) + v_{k,j}$$

其中, $h(\cdot)$ 表示路标在已知位姿变换的非线性模型, $z_{k,j}$ 表示路标 y_j 在当前帧的像素坐标, $v_{k,j} \sim N(0, Q_{k,j})$ 表示均值为0的,协方差为 $Q_{k,j}$ 的高斯噪声;

根据观测模型,依据投影位置 and 对应像素坐标形成的重投影误差建立误差模型:

$$e_{k,j} = z_{k,j} - h(\xi_k, y_j)$$

其中, $e_{k,j}$ 表示路标 y_j 在当前帧的位置和投影位置的差值, ξ_k 表示 k 时刻两帧之间位姿变换的李代数形;

将所述误差模型转化为非线性最小二乘问题,所有相机位姿 ξ 和路标 y 为待优化量 x ,追踪时间为 m ,路标总数为 n ,建立损失函数:

$$J(x) = \frac{1}{2} \sum_k^m \sum_j^n e_{k,j}(x)^T Q_{k,j}^{-1} e_{k,j}(x)$$

其中, $J()$ 表示损失函数;

通过解算所述损失函数得到优化位姿。

6. 根据权利要求5所述的基于光流和语义分割的视觉SLAM方法, 其特征在于, 所述步骤3中, 若输入的图像信息为双目相机对应的输入数据或深度相机对应的输入数据, 则对输入的图像信息中的特征点的类型进行判定并移除动态特征点包括以下子步骤:

采用重投影计算得到所述静态匹配点对集合所对应的第一偏移向量集合; 采用加权平均法, 根据所述静态区域 A_s 和所述第一偏移向量集合, 计算得到第一偏移向量权重 T_i 以及第一偏移向量权重的均值 ϕ_s ;

采用重投影计算得到所述预测运动匹配点对集合所对应的第二偏移向量集合; 采用加权平均法, 根据所述预测运动区域 A_m 和所述第二偏移向量集合, 计算得到第二偏移向量权重 T_j ;

根据所述第二偏移向量权重 T_j 和所述第一偏移向量权重的均值 ϕ_s , 对所述预测运动区域 A_m 中的每一个特征点的类型进行判断;

对预测运动区域是否为动态区域进行判断; 若预测运动区域中有超过第一阈值数量的特征点被判定为动态特征点, 则该预测运动区域标定为动态区域, 并去除标定为动态区域上的所有特征点。

7. 根据权利要求6所述的基于光流和语义分割的视觉SLAM方法, 其特征在于, 所述得到所述静态匹配点对集合所对应的第一偏移向量集合具体实现方式为:

所述步骤2中得到优化位姿对应的矩阵形式下旋转和平移分别为 R 、 t , 相机内参为 K , 设定前一帧和当前帧的匹配点对 p_i 和 $q_i(x_i, y_i)$ 对应三维空间点 P_i , 将 P_i 投影到当前帧得到投影坐标 $\hat{p}_i(\hat{x}_i, \hat{y}_i)$:

$$\hat{p}_i = K(RP_i + t)$$

其中, \hat{p}_i 表示将空间点 P_i 投影到当前帧的像素点位置, x_i 、 y_i 表示 q_i 的像素坐标, \hat{x}_i 、 \hat{y}_i 表示 \hat{p}_i 的像素坐标;

所述静态区域 A_s 中某一个匹配点对所对应的位置偏移向量表示为 $q_i \hat{p}_i = \vec{n}_i = (\hat{x}_i - x_i, \hat{y}_i - y_i)$; n 对匹配点对所对应的所述第一位置偏移向量集合表示为 $V_{state} = \{\vec{n}_i, i = 1, 2, 3, \dots, n\}$ 。

8. 根据权利要求7所述的基于光流和语义分割的视觉SLAM方法, 其特征在于, 所述得到第一偏移向量权重 T_i 、第一偏移向量权重的均值 ϕ_s 的具体实现方式为:

$$\theta_i = \tan^{-1}((\hat{y}_i - y_i)/(\hat{x}_i - x_i))$$

$$\varepsilon_i = \sqrt{(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2}$$

$$T_i = \frac{\varepsilon_i}{\sum_{k=1}^n \varepsilon_k} + \frac{|\theta_i|}{\sum_{k=1}^n |\theta_k|}, \quad i = 1, 2, 3, \dots, n$$

$$\phi_s = \frac{\sum_{i=1}^n T_i}{n}, \quad i = 1, 2, 3, \dots, n$$

其中, T_i 表示第一偏移向量 \vec{n}_i 在角度和模长上的权重, ϕ_s 表示第一偏移向量权重的均值;

所述对所述预测运动区域 A_m 中的每一个特征点的类型进行判断的具体实现方式为:

将 T_j 与 ϕ_s 进行对比,若 T_j 大于 ϕ_s ,则判定为动态特征点;否则,判定为静态特征点。

9. 根据权利要求5所述的基于光流和语义分割的视觉SLAM方法,其特征在于,所述步骤3中,若输入的图像信息为单目相机对应的输入数据,则对输入的图像信息中的特征点的类型进行判定并移除动态特征点包括以下子步骤:

根据在步骤2中得到优化位姿对应的旋转 R 和平移 t ,得到基础矩阵 F :

$$F = K^{-T} t \hat{R} K^{-1}$$

根据所述基础矩阵 F 、步骤2中得到的所述静态匹配点对集合和所述预测运动匹配点对集合,得到极线 $Fp_k = [x, y, z]^T$;

根据所述极线,得到特征点 D 值;

预设第二阈值 η ,根据所述特征点 D 值、所述第二阈值 η 判断特征点类型;

对预测运动区域是否为动态区域进行判断;若预测运动区域中有超过第一阈值数量的特征点被判定为动态特征点,则该预测运动区域标定为动态区域,并去除标定为动态区域上的所有特征点。

10. 根据权利要求9所述的基于光流和语义分割的视觉SLAM方法,其特征在于,所述特征点 D 值的计算方法为:

$$D = \frac{q_k^T F p_k}{\sqrt{\|x\|^2 + \|y\|^2}}, k = 1, 2, 3, \dots, n$$

其中, p_k, q_k 在公式中使用齐次形式, Fp_k 表示对极几何中的极线, x, y, z 为极线的向量参数表示, D 表示点 p_k 至极线 Fp_k 的距离;

若特征点 D 值大于第二阈值 η ,则判定点 p_k 为动态特征点;否则,判定为静态特征点。

一种基于光流和语义分割的视觉SLAM方法

技术领域

[0001] 本发明涉及视觉空间定位技术领域,尤其涉及一种基于光流和语义分割的视觉SLAM方法。

背景技术

[0002] SLAM是智能移动机器人领域中的一项关键技术,其中视觉SLAM就是使用相机作为主要传感器,相较其他类型的传感器,相机所能提供更多信息,所以近年来一直被广泛研究。但是,动态场景中实现准确的追踪和定位一直是SLAM系统的重大挑战。

[0003] 在实际场景中动态物体会在相机运动计算时造成错误数据,导致追踪失败或者错误追踪的情况。为解决该问题已经提出了几种方法,一种方法是传统的鲁棒性估计方法-RANSAC。这个方法会将动态信息等判定为外点剔除出去,保留静态信息保证追踪和运动计算成功,但是当环境中运动物体作为主体时造成提取的可用的数据太少,该方法就会失败。另一种方法集成额外的传感器。该方法可以利用多个传感器的数据信息互补实现补偿策略进行追踪和运动计算。但是,该方法设备成本和计算成本等并不经济,并且往往现实中该方法就是增加相机的数量。

[0004] 以往的方法应用到SLAM中并不算优秀,随着近年来,深度学习在语义分割,目标检测等方面的应用,对解决动态场景中运动物体的影响提供了一个新的解决方法。

[0005] 视觉SLAM可以分为两个类型,一种是基于特征的方法,一种是基于直接法。特征法通过比较特征的描述子来匹配点对和最小化重投影误差来实现追踪定位,该方法对于几何噪声能保持较好的鲁棒性。但是,提取特征点过程时间成本较大;直接法基于灰度不变理论通过计算重投影误差来优化位姿实现追踪,该方法比基于特征点在低纹理环境中性能更优,并且时间成本较低,但是,整体算法的鲁棒性较低。无论是特征点法还是直接法都不能解决常见的动态对象所带来的问题,并且这些动态对象会产生错误的关联降低计算位姿精度。

发明内容

[0006] 本申请实施例通过提供一种基于光流和语义分割的视觉SLAM方法,解决了动态环境中SLAM追踪定位效果较差的问题。

[0007] 本申请实施例提供一种基于光流和语义分割的视觉SLAM方法,包括以下步骤:

[0008] 步骤1、采用语义分割网络对输入的图像信息进行分割,得到静态区域、预测动态区域;

[0009] 步骤2、采用稀疏光流法对所述静态区域、所述预测动态区域进行特征追踪;

[0010] 步骤3、对输入的图像信息中的特征点的类型进行判定,并移除动态特征点;

[0011] 步骤4、将剔除运动特征点的集合作为追踪数据,输入至ORB-SLAM中进行处理,输出位姿结果。

[0012] 优选的,所述步骤1中的所述输入的图像信息为单目相机对应的输入数据、双目相

机对应的输入数据、深度相机对应的输入数据中的一种；

[0013] 所述步骤3中,通过对极约束对单目相机对应的输入数据中的特征点的类型进行判断;通过重投影误差对双目相机对应的输入数据或深度相机对应的输入数据中的特征点的类型进行判断。

[0014] 优选的,所述步骤1包括以下子步骤:

[0015] 步骤1.1、选取数据集对Mask R-CNN网络进行训练,得到训练好的语义分割网络;所述数据集包括多种类型数据作为潜在的运动对象;

[0016] 步骤1.2、将图像信息输入至训练好的语义分割网络,完成图像分割,得到静态区域 A_s 、预测运动区域 A_m 。

[0017] 优选的,所述步骤2包括以下子步骤:

[0018] 步骤2.1、利用稀疏光流法对所述静态区域 A_s 和所述预测运动区域 A_m 进行特征提取、匹配,得到静态匹配点对集合和预测运动匹配点对集合;

[0019] 步骤2.2、基于SLAM的运行模型求解位姿。

[0020] 优选的,所述步骤2.2中的所述基于SLAM的运行模型求解位姿包括:

[0021] 在 k 时刻,利用第 j 个路标 y_j 投影到当前帧得到投影位置 $h(\xi_k, y_j)$,得到对应的观测模型:

$$[0022] \quad z_{k,j} = h(\xi_k, y_j) + v_{k,j}$$

[0023] 其中, $h(\cdot)$ 表示路标在已知位姿变换的非线性模型, $z_{k,j}$ 表示路标 y_j 在当前帧的像素坐标, $v_{k,j} \sim N(0, Q_{k,j})$ 表示均值为0的,协方差为 $Q_{k,j}$ 的高斯噪声;

[0024] 根据观测模型,依据投影位置 and 对应像素坐标形成的重投影误差建立误差模型:

$$[0025] \quad e_{k,j} = z_{k,j} - h(\xi_k, y_j)$$

[0026] 其中, $e_{k,j}$ 表示路标 y_j 在当前帧的位置和投影位置的差值, ξ_k 表示 k 时刻两帧之间位姿变换的李代数形;

[0027] 将所述误差模型转化为非线性最小二乘问题,所有相机位姿 ξ 和路标 y 为待优化量 x ,追踪时间为 m ,路标总数为 n ,建立损失函数:

$$[0028] \quad J(x) = \frac{1}{2} \sum_k^m \sum_j^n e_{k,j}(x)^T Q_{k,j}^{-1} e_{k,j}(x)$$

[0029] 其中, $J(\cdot)$ 表示损失函数;

[0030] 通过解算所述损失函数得到优化位姿。

[0031] 优选的,所述步骤3中,若输入的图像信息为双目相机对应的输入数据或深度相机对应的输入数据,则对输入的图像信息中的特征点的类型进行判定并移除动态特征点包括以下子步骤:

[0032] 采用重投影计算得到所述静态匹配点对集合所对应的第一偏移向量集合;采用加权平均法,根据所述静态区域 A_s 和所述第一偏移向量集合,计算得到第一偏移向量权重 T_i 以及第一偏移向量权重的均值 ϕ_s ;

[0033] 采用重投影计算得到所述预测运动匹配点对集合所对应的第二偏移向量集合;采用加权平均法,根据所述预测运动区域 A_m 和所述第二偏移向量集合,计算得到第二偏移向量权重 T_j ;

[0034] 根据所述第二偏移向量权重 T_j 和所述第一偏移向量权重的均值 ϕ_s ,对所述预测运动区域 A_m 中的每一个特征点的类型进行判断;

[0035] 对预测运动区域是否为动态区域进行判断;若预测运动区域中有超过第一阈值数量的特征点被判定为动态特征点,则该预测运动区域标定为动态区域,并去除标定为动态区域上的所有特征点。

[0036] 优选的,所述得到所述静态匹配点对集合所对应的第一偏移向量集合具体实现方式为:

[0037] 所述步骤2中得到优化位姿对应的矩阵形式下旋转和平移分别为 R 、 t ,相机内参为 K ,设定前一帧和当前帧的匹配点对 p_i 和 $q_i(x_i, y_i)$ 对应三维空间点 P_i ,将 P_i 投影到当前帧得到投影坐标 $\hat{p}_i(\hat{x}_i, \hat{y}_i)$:

$$[0038] \quad \hat{p}_i = K(RP_i + t)$$

[0039] 其中, \hat{p}_i 表示将空间点 P_i 投影到当前帧的像素点位置, x_i, y_i 表示 q_i 的像素坐标, \hat{x}_i, \hat{y}_i 表示 \hat{p}_i 的像素坐标;

[0040] 所述静态区域 A_s 中某一个匹配点对所对应的位置偏移向量表示为 $q_i\hat{p}_i = \vec{n}_i = (\hat{x}_i - x_i, \hat{y}_i - y_i)$; n 对匹配点对所对应的所述第一位置偏移向量集合表示为 $V_{state} = \{\vec{n}_i, i = 1, 2, 3, \dots, n\}$ 。

[0041] 优选的,所述得到第一偏移向量权重 T_i 、第一偏移向量权重的均值 ϕ_s 的具体实现方式为:

$$[0042] \quad \theta_i = \tan^{-1}((\hat{y}_i - y_i)/(\hat{x}_i - x_i))$$

$$[0043] \quad \varepsilon_i = \sqrt{(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2}$$

$$[0044] \quad T_i = \frac{\varepsilon_i}{\sum_{k=1}^n \varepsilon_k} + \frac{|\theta_i|}{\sum_{k=1}^n |\theta_k|}, \quad i = 1, 2, 3, \dots, n$$

$$[0045] \quad \phi_s = \frac{\sum_{i=1}^n T_i}{n}, \quad i = 1, 2, 3, \dots, n$$

[0046] 其中, T_i 表示第一偏移向量 \vec{n}_i 在角度和模长上的权重, ϕ_s 表示第一偏移向量权重的均值;

[0047] 所述对所述预测运动区域 A_m 中的每一个特征点的类型进行判断的具体实现方式为:

[0048] 将 T_j 与 ϕ_s 进行对比,若 T_j 大于 ϕ_s ,则判定为动态特征点;否则,判定为静态特征点。

[0049] 优选的,所述步骤3中,若输入的图像信息为单目相机对应的输入数据,则对输入的图像信息中的特征点的类型进行判定并移除动态特征点包括以下子步骤:

[0050] 根据在步骤2中得到优化位姿对应的旋转 R 和平移 t ,得到基础矩阵 F :

$$[0051] \quad F = K^{-T} t \hat{R} K^{-1}$$

[0052] 根据所述基础矩阵 F 、步骤2中得到的所述静态匹配点对集合和所述预测运动匹配

点对集合,得到极线 $Fp_k = [x, y, z]^T$;

[0053] 根据所述极线,得到特征点D值;

[0054] 预设第二阈值 η ,根据所述特征点D值、所述第二阈值 η 判断特征点类型;

[0055] 对预测运动区域是否为动态区域进行判断;若预测运动区域中有超过第一阈值数量的特征点被判定为动态特征点,则该预测运动区域标定为动态区域,并去除标定为动态区域上的所有特征点。

[0056] 优选的,所述特征点D值的计算方法为:

$$[0057] \quad D = \frac{q_k^T F p_k}{\sqrt{\|x\|^2 + \|y\|^2}}, k = 1, 2, 3, \dots, n$$

[0058] 其中, p_k, q_k 在公式中使用齐次形式, Fp_k 表示对极几何中的极线, x, y, z 为极线的向量参数表示, D 表示点 p_k 至极线 Fp_k 的距离;

[0059] 若特征点D值大于第二阈值 η ,则判定点 p_k 为动态特征点;否则,判定为静态特征点。

[0060] 本申请实施例中提供的一个或多个技术方案,至少具有如下技术效果或优点:

[0061] 在本申请实施例中,首先采用语义分割网络对输入的图像信息进行分割,得到静态区域、预测动态区域,然后采用稀疏光流法对静态区域、预测动态区域进行特征追踪,之后对输入的图像信息中的特征点的类型进行判定,并移除动态特征点,最后将剔除运动特征点的集合作为追踪数据,输入至ORB-SLAM中进行处理,输出位姿结果。即本发明基于语义分割和光流法结合的动态对象提取、判定并移除动态影响,将去除动态影响后的静态特征点应用到后续的SLAM系统中,最终可以得到动态环境中位姿精度较高的轨迹信息。相对于传统的对动态环境处理方法,本发明可以很好地判定并消除动态物体的特征影响和位姿精度不高的问题。

附图说明

[0062] 为了更清楚地说明本实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图是本发明的一个实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0063] 图1为本发明实施例提供的一种基于光流和语义分割的视觉SLAM方法的整体流程图。

具体实施方式

[0064] 本实施例提供一种基于光流和语义分割的视觉SLAM方法,主要包括以下步骤:

[0065] 步骤1、采用语义分割网络对输入的图像信息进行分割,得到静态区域、预测动态区域。

[0066] 步骤2、采用稀疏光流法对所述静态区域、所述预测动态区域进行特征追踪。

[0067] 步骤3、对输入的图像信息中的特征点的类型进行判定,并移除动态特征点。

[0068] 步骤4、将剔除运动特征点的集合作为追踪数据,输入至ORB-SLAM中进行处理,输出位姿结果。

[0069] 为了更好的理解上述技术方案,下面将结合说明书附图以及具体的实施方式对上

述技术方案进行详细的说明。

[0070] 本实施例提供了一种基于光流和语义分割的视觉SLAM方法,如图1所示,包括以下步骤:

[0071] 步骤1、使用Mask R-CNN网络对输入的图像信息(数据)进行分割,区分静态和动态的对象,得到静态区域、预测动态区域。

[0072] 步骤1.1、选取数据集对Mask R-CNN网络进行训练,在COCO数据集中选取20种类型数据作为潜在的运动对象,如:人,自行车,公共汽车,船,鸟,猫,狗等。

[0073] 步骤1.2、读取单目相机、双目相机或者深度相机RGB-D对应的数据输入网络。在训练完成的语义分割网络中输入图像的格式是 $m \times n \times 3$,输出结果的格式为 $m \times n \times 1$,其中 $m \times n$ 表示图像的大小,3表示图像通道(RGB),1代表在步骤1.1中选取的训练类别数目(即20),完成语义分割,结合选取的20种类型作为可能移动的物体并标记区域,完成图像分割得到静态区域 A_s 和预测运动区域 A_m 。如果分割未得到 A_m ,此时就可认为数据中不存在动态区域,没有必要进行动态区域的处理,可以按照传统的ORB-SLAM进行处理,直接跳转到步骤4。

[0074] 步骤2、采用稀疏光流法对所述静态区域、所述预测动态区域进行特征追踪。

[0075] 完成数据的预处理之后,利用轻量级算法进行追踪,算法本质即在ORB-SLAM追踪的基础上保留特征提取和追踪功能,其中采用光流法替代特征点法,并剔除了局部优化和关键帧决策的功能模块,完成特征提取,匹配和位姿求解。

[0076] 步骤2.1、对于已经完成分割的数据,利用Lucas-Kanade光流法对于静态区域 A_s 和预测运动区域 A_m 进行特征提取和匹配,对当前帧得到所有特征点集合 P ,匹配得到静态匹配点对集合 $P_{\text{match}s} = \{(p_i, q_i), i = 1, 2, 3, \dots, n\}$ 和预测运动匹配点对集合 $P_{\text{match}m} = \{(p_j, q_j), j = 1, 2, 3, \dots, m\}$ 。 p_i 和 q_i 分别表示前一帧和当前帧的第 i 个静态匹配像素点对, p_j 和 q_j 分别表示前一帧和当前帧的第 j 个预测运动匹配像素点对, n 、 m 分别表示静态区域匹配点对的数目、预测运动区域匹配点对的数目。

[0077] 步骤2.2、基于SLAM的运行模型开始求解位姿,此时在 k 时刻利用第 j 个路标 y_j 投影到当前帧得到投影位置 $h(\xi_k, y_j)$,此时可以得到对应的观测模型:

$$[0078] \quad z_{k,j} = h(\xi_k, y_j) + v_{k,j}$$

[0079] 其中, $h(\cdot)$ 表示路标在已知位姿变换的非线性模型, $z_{k,j}$ 表示路标 y_j 在当前帧像素坐标, $v_{k,j} \sim N(0, Q_{k,j})$ 表示均值为0的,协方差为 $Q_{k,j}$ 的高斯噪声。

[0080] 根据观测模型可以依据投影位置和对应该像素坐标形成的重投影误差建立误差模型:

$$[0081] \quad e_{k,j} = z_{k,j} - h(\xi_k, y_j)$$

[0082] 其中, $e_{k,j}$ 表示路标 y_j 在当前帧的位置和投影位置的差值, ξ_k 表示 k 时刻两帧之间位姿变换的李代数形。

[0083] 步骤2.3、将误差模型转化为非线性最小二乘问题,所有相机位姿 ξ 和路标 y 为待优化量 x ,追踪时间为 m ,路标总数为 n ,建立如下损失函数:

$$[0084] \quad J(x) = \frac{1}{2} \sum_k^m \sum_j^n e_{k,j}(x)^T Q_{k,j}^{-1} e_{k,j}(x)$$

[0085] 其中, $J(\cdot)$ 表示损失函数, k 表示追踪的 k 时刻, j 表示第 j 个路标, $e_{k,j}$ 表示步骤2.2中

的误差, $Q_{k,j}$ 表示高斯噪声的协方差。

[0086] 通过解算损失函数得出相机的优化位姿。

[0087] 步骤3、对输入的图像信息中的特征点的类型进行判定,并移除动态特征点。

[0088] 输入的图像信息为单目相机对应的输入数据、双目相机对应的输入数据、深度相机对应的输入数据中的一种。依据不同的传感器类型对应不同的输入数据类型,设计不同的动态特征点判断和处理方法。

[0089] 其中,深度相机RGB-D和双目相机类型使用相同方法处理,进入步骤3.1;单目相机类型则直接跳转至步骤3.4。

[0090] 步骤3.1、本环节针对的是RGB-D和双目系统,在步骤2中得到相机的优化后的初始位姿李代数形式为 ξ ,对应的矩阵形式下旋转和平移为 R 、 t ,已知相机内参 K ,设定前一帧和当前帧的匹配像素点对 p_i 和 $q_i(x_i, y_i)$ 对应三维空间点 P_i ,将 P_i 投影到当前帧得到投影坐标 $\hat{p}_i(\hat{x}_i, \hat{y}_i)$,关系如下:

$$[0091] \quad \hat{p}_i = K(RP_i + t)$$

[0092] 其中, \hat{p}_i 表示将空间点 P_i 投影到当前帧的像素点位置, x_i 、 y_i 表示 q_i 的像素坐标; \hat{x}_i 、 \hat{y}_i 表示 \hat{p}_i 的像素坐标。

[0093] 如果不存在误差影响,应该有 $\hat{p}_i = q_i$ 。但是,由于噪声等影响,静态特征点和预测运动点会存在的位置偏移造成 $\hat{p}_i \neq q_i$ 。

[0094] 其中,某一静态匹配点对所对应的位置偏移向量表示为 $q_i \hat{p}_i = \vec{n}_i = (\hat{x}_i - x_i, \hat{y}_i - y_i)$,此时,对于 n 对匹配点对则存在偏移向量的集合 $V_{state} = \{\vec{n}_i, i = 1, 2, 3, \dots, n\}$,即静态匹配点对集合所对应的第一偏移向量集合。

[0095] 步骤3.2、根据步骤2.1中得到的静态匹配点对,使用加权平均法描述静态区域 A_s 的平移量,针对静态区域 A_s 和第一偏移向量集合 V_{state} ,计算偏移误差 \vec{n}_i 的角度 θ_i 、模长 ε_i 和权重 T_i :

$$[0096] \quad \theta_i = \tan^{-1}((\hat{y}_i - y_i)/(\hat{x}_i - x_i))$$

$$[0097] \quad \varepsilon_i = \sqrt{(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2}$$

$$[0098] \quad T_i = \frac{\varepsilon_i}{\sum_{k=1}^n \varepsilon_k} + \frac{|\theta_i|}{\sum_{k=1}^n |\theta_k|}, \quad i = 1, 2, 3, \dots, n$$

[0099] 其中, T_i 表示偏移向量 \vec{n}_i 在角度和模长上的权重。

[0100] 然后,计算偏移向量权重的均值 ϕ_s :

$$[0101] \quad \phi_s = \frac{\sum_{i=1}^n T_i}{n}, \quad i = 1, 2, 3, \dots, n$$

[0102] 步骤3.3、采用重投影计算得到预测运动匹配点对集合所对应的第二偏移向量集合;采用加权平均法,根据预测运动区域 A_m 和第二偏移向量集合,计算得到第二偏移向量权重 T_j 。

[0103] 即参照步骤3.1和3.2获取对应预测运动区域 A_m 的第二偏移向量集合

$V_{other} = \{\overrightarrow{m}_j, j = 1, 2, 3, \dots, m\}$, 以及偏移误差的角度、模长和对应的第二偏移向量权重 T_j , 将 T_j 与 ϕ_s 对比:

$$[0104] \quad \begin{cases} T_j > \phi_s, p_j \text{ 是动态特征点} \\ T_j < \phi_s, p_j \text{ 是静态特征点} \end{cases}$$

[0105] 利用 V_{other} 中的每个向量完成对应的预测特征点动态或静态的判定。完成此步骤后跳转至步骤3.6。

[0106] 步骤3.4、此步骤针对单目系统,在步骤2中得到相机的优化后的初始位姿李代数形为 ξ , ξ 对应的旋转和平移为 R 、 t ,此时得到本次运动的基础矩阵 F :

$$[0107] \quad F = K^{-T} t \hat{R} K^{-1}$$

[0108] 同时,步骤2.1中得到两帧图像特征匹配的所有点对(包含静态区域匹配点对和预测运动区域匹配点对)的集合 $P_{match} = \{(p_k, q_k), k = 1, 2, 3, \dots, n\}$, n 代表所有匹配点对总数,结合基础矩阵 F ,得到极线 $Fp_k = [x, y, z]^T$ 。

$$[0109] \quad D = \frac{q_k^T F p_k}{\sqrt{\|x\|^2 + \|y\|^2}}, k = 1, 2, 3, \dots, n$$

[0110] 其中, p_k, q_k 在公式中使用齐次形式, Fp_k 表示对极几何中的极线, x, y, z 为极线的向量参数表示, D 表示点 p_k 至极线 Fp_k 的距离。

[0111] 步骤3.5、集合 P 中点对依次计算特征点 D 值,设定阈值 η ,根据特征点 D 值和阈值 η 判断特征点类型:

$$[0112] \quad \begin{cases} D > \eta, p_k \text{ 是动态特征点} \\ D < \eta, p_k \text{ 是静态特征点} \end{cases}$$

[0113] 其中,在测试中发现 η 设定为5的效果较为稳定,所以优选地采用 $\eta = 5$ 来解算。

[0114] 步骤3.6、依据动态特征点集合,对步骤1.2中提取的预测运动区域进行判断,当确定预测运动区中大部分特征点(例如,80%以上)都判定为动态特征点,则判定该区域为动态区域,之后去除标定为动态区域上的所有特征点。

[0115] 步骤4、将所有的特征点(包含静态区域和预测动态区域对应的特征点)中剔除运动特征点的集合 P_e 作为追踪数据,此时集合 P_e 就去除了动态特征的影响,之后输入传统的ORB-SLAM的框架中进行处理,输出位姿结果。

[0116] 步骤4.1、集合 P_e 进入追踪中完成局部地图创建,位姿优化。

[0117] 步骤4.2、进行闭环检测。

[0118] 步骤4.3、输出位姿结果。

[0119] 综上,本发明提出的一种基于语义分割和光流法结合的动态对象提取、判定并移除动态影响的SLAM方法,该方法采用语义分割网络可以有效地分割出潜在的动态对象,接着使用稀疏光流法完成稳定的特征追踪,然后通过匹配点对的对极约束以及重投影误差分布差异判定并移除动态特征点,将去除动态影响后的静态特征点应用到后续的SLAM系统中,最终可以得到动态环境中位姿精度较高的轨迹信息。相对于传统的对动态环境处理方

法,本发明可以很好地判定并消除动态物体的特征影响和位姿精度不高的问题。

[0120] 最后所应说明的是,以上具体实施方式仅用以说明本发明的技术方案而非限制,尽管参照实例对本发明进行了详细说明,本领域的普通技术人员应当理解,可以对本发明的技术方案进行修改或者等同替换,而不脱离本发明技术方案的精神和范围,其均应涵盖在本发明的权利要求范围当中。

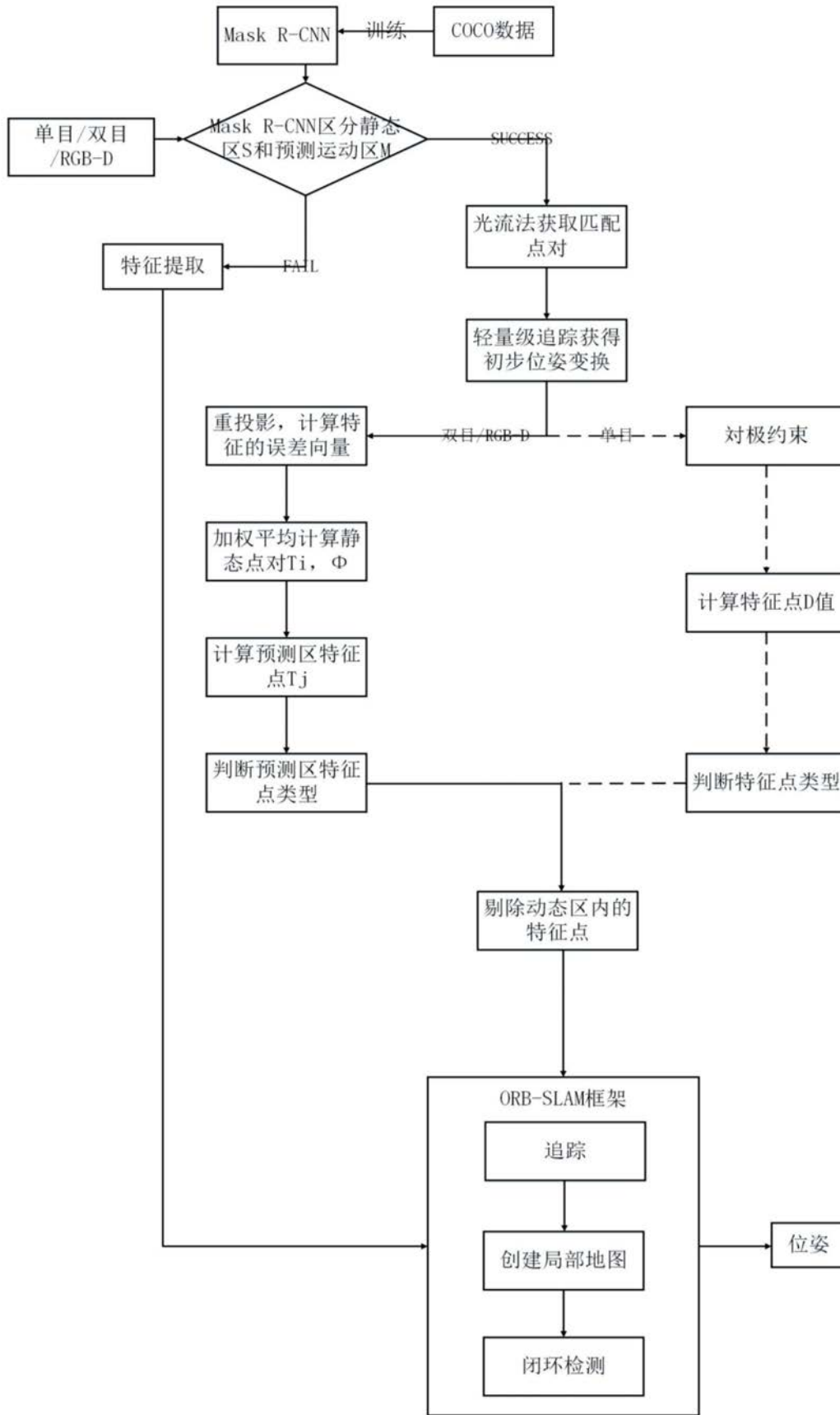


图1