



(12) 发明专利申请

(10) 申请公布号 CN 111798475 A

(43) 申请公布日 2020.10.20

(21) 申请号 202010477740.X

G06F 16/29 (2019.01)

(22) 申请日 2020.05.29

(71) 申请人 浙江工业大学

地址 310014 浙江省杭州市下城区潮王路18号

(72) 发明人 朱威 绳荣金 陈璐瑶 郑雅羽 何德峰

(74) 专利代理机构 杭州赛科专利代理事务所 (普通合伙) 33230

代理人 郭薇

(51) Int. Cl.

G06T 7/13 (2017.01)

G06T 7/11 (2017.01)

G06N 3/04 (2006.01)

G06K 9/62 (2006.01)

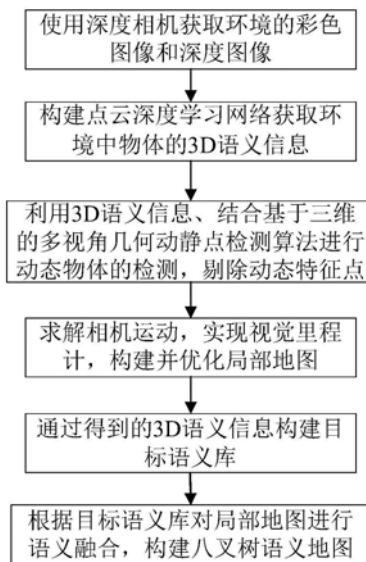
权利要求书2页 说明书9页 附图8页

(54) 发明名称

一种基于点云深度学习的室内环境3D语义地图构建方法

(57) 摘要

本发明涉及一种基于点云深度学习的室内环境3D语义地图构建方法,包括以下四部分:(1)使用深度相机获取室内环境的彩色图和深度图;(2)构建点云深度学习网络获取环境中物体的3D语义信息;(3)动态物体的检测与动态特征点的剔除;(4)求解相机运动实现视觉里程计,构建并优化局部地图;(5)通过得到的3D点云语义信息构建目标语义库;(6)根据目标语义库对局部地图进行语义融合,构建八叉树语义地图。与现有方法相比,本发明不仅通过结合语义类别信息剔除了动态物体掩码中的特征点,有效减少了动态物体对定位和建图的影响,而且使用的3D语义获取方式更为直接高效,具有更好的定位建图能力以及语义感知效果。



1. 一种基于点云深度学习的室内环境3D语义地图构建方法,其特征在于:所述方法包括以下步骤:

步骤1:使用深度相机获取环境的彩色图像和深度图像;

步骤2:构建点云深度学习网络获取环境中物体的3D语义信息;

步骤3:利用3D语义信息、结合基于三维的多视角几何动静点检测算法进行动态物体的检测,剔除动态特征点;

步骤4:求解相机运动,实现视觉里程计,构建并优化局部地图;

步骤5:通过得到的3D语义信息构建目标语义库;

步骤6:根据目标语义库对局部地图进行语义融合,构建八叉树语义地图。

2. 根据权利要求1所述的一种基于点云深度学习的室内环境3D语义地图构建方法,其特征在于:所述步骤2包括以下子步骤:

步骤2.1:将点云深度学习网络PointNet中通过多层感知器对点云进行特征提取的方法改为通过构造动态图卷积提取特征的方法;

所构造的动态图卷积是基于空间的图卷积,通过选取k个近邻点把点云构造为有边和顶点的图,k的取值范围为[10,50],然后在图上进行卷积,学习选取的顶点与其相邻点之间组成的边的关系来获得点云局部特征;

用D表示点云的维度,对于一帧点数为n、维度为D的点云,可以表示为式(1), $R^D$ 表示D维度的点云在实数域的集合;

$$X = \{x_1, \dots, x_n\} \subseteq R^D \quad (1)$$

点云中的第i个点 $x_i$ 可以分别用式(2)-(4)表示,其中当D=3时,表示网络输入具有u、v、w三个坐标维度的点云,用式(2)表示;D=6时,表示网络输入带有r、g、b信息的彩色点云,用式(3)表示;D=9表示输入带有曲面法向量 $\alpha, \beta, \gamma$ 的点云,用式(4)表示;

$$x_i = (u_i, v_i, w_i) \quad (2)$$

$$x_i = (u_i, v_i, w_i, r_i, g_i, b_i) \quad (3)$$

$$x_i = (u_i, v_i, w_i, r_i, g_i, b_i, \alpha_i, \beta_i, \gamma_i) \quad (4)$$

从第1层的D维度的点云特征中选取的第i个顶点 $x_i^{(l)}$ 与其 $k_1$ 个近邻点经过图卷积学习边缘特征,并经过聚合操作 $\Sigma$ ,得到的第(1+1)层的D维度的特征输出 $x_i^{(l+1)}$ 用式(5)表示;其中, $x_j^{(l)}$ 表示第i个顶点 $x_i^{(l)}$ 的近邻点, $\epsilon^{(l)}$ 代表 $x_i^{(l)}$ 和 $x_j^{(l)}$ 构成的顶点和近邻点的集合, $h_{\Theta}$ 是被可学习的参数 $\Theta$ 集合参数化的一系列非线性函数, $h_{\Theta}^{(l)}$ 表示在第1层上的特征学习过程 $R^{D(1)} \times R^{D(1)} \rightarrow R^{D(1+1)}$ ;

$$x_i^{(l+1)} = \sum_{j:(i,j) \in \epsilon^{(l)}} h_{\Theta}^{(l)}(x_i^{(l)}, x_j^{(l)}) \quad (5)$$

步骤2.2:使用多窗口和窗口步长的空间金字塔池化结构对学习的点云特征分空间层次地进行多尺度池化;空间金字塔池化结构式如式(6)所示,

$$F(x_1, \dots, x_N) = \underset{x=x_1, \dots, x_N}{mlp} [c\{g(f, s_1), \dots, g(f, s_n)\}] \quad (6)$$

空间金字塔池化窗口类型为N、N/4、N/8、N/16的组合,其中N为输入的点云个数,f代表

通过图卷积网络学习得到的特征信息,  $s_n$ 表示第 $n$ 个池化窗口的尺寸,  $g$ 代表池化操作, 多个不同的池化窗口组合构成了空间金子塔池化结构,  $c$ 代表将空间金字塔池化得到的多尺度特征进行聚合, 经过多层全连接神经网络mlp进行上采样, 得到点云中 $N$ 个点的每个点的特征 $F(x_1, \dots, x_N)$ 。

3. 根据权利要求1所述的一种基于点云深度学习的室内环境3D语义地图构建方法, 其特征在于: 所述步骤3包括以下子步骤:

步骤3.1: 利用点云深度学习网络获取的环境点云语义信息预筛选可能存在的先验动态物体;

步骤3.2: 使用基于三维的多视角几何动静点检测方法检测是否存在动态的特征点; 将构建的局部地图中的点投影到当前相机视角下, 测量当前帧深度值 $d'$ 和投影的深度值 $d$ 的差值绝对值 $diff$ , 若 $diff$ 大于阈值 $TH$ , 说明是有动态物体遮挡了当前场景的部分视野, 则判定被遮挡场景前方物体即为动态物体; 室内环境中阈值 $TH$ 取值范围为 $[0.3, 1]$ , 单位米;

步骤3.3: 输出动态物体信息, 剔除动态物体对应的动态特征点;

所述动态物体信息包括区域位置和对应的特征点信息。

4. 根据权利要求1所述的一种基于点云深度学习的室内环境3D语义地图构建方法, 其特征在于: 所述步骤4包括以下子步骤:

步骤4.1: 对已剔除动态特征点的参考帧和当前帧进行特征匹配, 估计相机位姿, 跟踪视觉里程计并生成局部地图, 选取新的关键帧;

步骤4.2: 基于ORB-SLAM2算法, 插入新关键帧、剔除冗余地图点、更新创建新地图点, 利用光束平差法进行局部地图的相机位姿优化, 构建更新局部地图。

5. 根据权利要求1所述的一种基于点云深度学习的室内环境3D语义地图构建方法, 其特征在于: 所述步骤5包括以下子步骤:

步骤5.1: 通过语义分割得到的当前点云帧的语义信息, 初始化构建包含RGB-D图像、3D包围框、ORB特征点、物体类别和类别概率的语义库;

步骤5.2: 对当前点云关键帧进行语义分割, 并在语义库中按照物体类别查找是否存在同类别的语义信息;

步骤5.3: 若语义库中没有与目标物体对应的物体类别信息, 则将目标物体添加到语义库中; 若在超过连续若干帧中检测到相同的目标语义类别, 则计算新检测到的目标物体3D包围框中心坐标与语义库已有同类别所有物体3D包围框中心坐标的欧氏距离, 找到距离最小的物体中心坐标, 根据位置信息判断是否是同一物体; 若是同一物体, 则更新该物体的空间坐标信息, 若不是同一个物体, 则将此新物体加入语义库。

6. 根据权利要求5所述的一种基于点云深度学习的室内环境3D语义地图构建方法, 其特征在于: 所述步骤6包括以下子步骤:

步骤6.1: 在步骤4构建的局部地图基础上, 根据贝叶斯法则更新融合深度相机观测数据, 进行三维占有格概率动态更新, 利用步骤5得到的目标语义库不断更新对应位置八叉树节点的语义信息, 实现八叉树地图的语义化;

步骤6.2: 通过局部优化建图模板对构建的八叉树地图进行局部地图的相机位姿优化和地图配准对齐; 然后使用闭环优化模块对全局地图进行相机位姿和空间点位置的集束调整优化。

## 一种基于点云深度学习的室内环境3D语义地图构建方法

### 技术领域

[0001] 本发明属于计算机视觉技术在移动机器人领域的应用,具体涉及一种基于点云深度学习的室内环境3D语义地图构建方法。

### 背景技术

[0002] 同时定位和地图构建(Simultaneous Localization and Mapping,SLAM)算法是基于传感器数据信息同时实现传感器自身位姿估计和环境地图构建的过程,是移动机器人和无人驾驶汽车在未知环境中实现完全自主导航的关键技术。依照传感器的种类不同,SLAM分为基于激光的SLAM和基于视觉的SLAM两大类,其中基于激光传感器的SLAM研究较早,理论和技术相对成熟,然而高精度的激光传感器价格昂贵,性价比较低。以摄像机作为传感器的SLAM方案被称作视觉SLAM(Vision-based SLAM,VSLAM)。传统的视觉SLAM重在几何的定位和建图,具有高性价比的优势,但还没有感知环境的语义信息。因此,在其基础上,能够构建具有语义信息地图的语义SLAM是目前工业界重要研发方向。

[0003] 语义SLAM的研发尚处于起步阶段,目前发表的论文和公开的专利主要是基于2D图像的目标检测或语义分割来得到环境的语义信息,然后将获得的语义信息映射到聚类分割出的3D点云团,从而间接得到对应帧的3D点云语义信息。这种获取目标语义信息的方式不是基于对原生3D点云的直接特征提取,存在处理方式不直接、语义效果不佳、深度信息不充分、3D边缘分割模糊等问题。另外,传统的视觉SLAM和现有语义SLAM方案主要局限于对静态环境的处理,无法很好地应用于复杂动态场景。

[0004] Sünderhauf等人于2016年提出了面向物体对象的语义建图方法,利用ORB-SLAM2和SSD通过最近邻的方法进行语义数据融合来实现语义地图的构建(见Sünderhauf N,Pham T,Latif Y,et al.Meaningful maps Object-oriented semantic mapping[C].//International Conference on Intelligent Robots and Systems.USA:IEEE,2016:1609-7849)。此外,还有研究人员利用目标检测算法YOLO结合ORB-SLAM算法构建语义地图(见白云汉.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(01):183-190)。以上语义地图构建方法都依赖于在2D的图像数据上用卷积神经网络进行语义分割,并非对深度传感器得到的原始3D点云数据进行直接学习然后融合到SLAM系统中实现3D语义地图构建,所以构建出的语义地图对于环境信息的表达能力是有限的,在室内建图导航等应用场景中的应用还是有一定的局限性。

[0005] 申请号为201811089153.2的专利公开了一种识别目标的坐标系定位和语义地图构建的方法及系统。该方法通过对作为目标物体的特征物体,将其目标检测2D框与采集的对应点云数据融合,对目标物体在世界坐标系下进行三维定位,基于上述定位方法进行栅格-语义地图的构建。但该方法仍然依赖2D语义分割网络来得到3D点云语义信息,且只适用于静态环境下的栅格语义地图构建,若存在动态物体,则会极大的影响地图构建的精度。申请号为201910408713.4的专利公开了一种基于深度学习的室内环境3D语义地图构建方法。该发明用训练好的基于RGB-D图像的语义分割网络对当前获取的RGB-D图像进行特征提取

与处理,然后用关键帧构建点云地图,并将新获取图像帧对应点云与已构建的点云地图进行融合,将像关键帧的语义标注结果映射到对应点云地图上构建3D语义地图。该发明在语义获取方式上也是基于2D的语义分割网络,且适用于静态环境的语义地图构建,另外构建的3D语义地图是点云形式的语义地图,非拓扑结构的点云语义地图不能直接用于机器人在室内的定位导航和避障等功能。

[0006] Qi等人提出的点云深度学习网络PointNet(见Qi C R,Su H,Kaichun M,etal.PointNet:deep learning on point sets for 3D classification and segmentation[C].//International Conference on Computer Vision and Pattern Recognition.Honolulu:IEEE Computer Society Press,2017:77-85.)。开创性的直接在非结构化的点云数据上进行特征学习,得到3D点云的语义信息,为直接利用3D语义信息融合构建室内环境语义地图提供了技术可能。

### 发明内容

[0007] 为了解决现有视觉SLAM地图构建方法存在的对动态场景适应性差、无语义感知能力或语义获取方式复杂不直接、构建出的点云地图占用存储空间巨大、非拓扑地图结构不能够直接用于导航避障等问题,本发明提供了一种基于点云深度学习的室内环境3D语义地图构建方法,所述方法包括以下步骤:

[0008] (1) 使用深度相机获取室内环境的彩色图像和深度图像;

[0009] (2) 构建点云深度学习网络获取环境中物体的3D语义信息;

[0010] 构建点云深度学习网络,对从深度相机获取的彩色图像帧对应的点云帧进行语义分割,得到环境的点云语义信息。所述的点云深度学习网络设计如下:

[0011] (2-1) 将点云深度学习网络PointNet中通过多层感知器(Multi-Layer Perceptron,MLP)模块对点云进行特征提取的方法改为通过构造动态图卷积模块提取特征的方法。

[0012] 把数据构造成有边和顶点的图,然后在图上进行卷积的操作被称为图卷积。本发明的图卷积是基于空间的图卷积,通过选取k个近邻点把点云数构造为有边和顶点的图,k取值范围为[10,50],然后在图上进行卷积,学习选取的顶点与其相邻点之间组成的边的关系来获得点云局部特征。图卷积是动态不固定的,一个点的k个近邻点的选取对象随着网络层和特征的不同而变化,使得每一层图的参数也随着层的不同而动态更新,从而构成了动态的图卷积网络。

[0013] 点云的维度可以是3维、6维、9维等,用D表示点云的维度。对于一帧点数为n、维度为D的点云,可以表示为式(1), $R^D$ 表示D维度的点云在实数域的集合。

$$[0014] \quad X = \{x_1, \dots, x_n\} \subseteq R^D \quad (1)$$

[0015] 点云中的第i个点 $x_i$ 可以分别用式(2)-(4)表示,其中当D=3时,表示网络输入具有u、v、w三个坐标维度的点云,用式(2)表示;D=6时,表示网络输入带有r、g、b信息的彩色点云,用式(3)表示;D=9表示输入带有曲面法向量 $\alpha, \beta, \gamma$ 的点云,用式(4)表示。

$$[0016] \quad x_i = (u_i, v_i, w_i) \quad (2)$$

$$[0017] \quad x_i = (u_i, v_i, w_i, r_i, g_i, b_i) \quad (3)$$

$$[0018] \quad x_i = (u_i, v_i, w_i, r_i, g_i, b_i, \alpha_i, \beta_i, \gamma_i) \quad (4)$$

[0019] 从第1层的D维度的点云特征中选取的第i个顶点  $x_i^{(l)}$  与其  $k_1$  个近邻点经过图卷积学习边缘特征,并经过聚合操作  $\Sigma$ ,得到的第 (1+1) 层的D维度的特征输出  $x_i^{(l+1)}$  可以用式 (5) 表示。其中,  $x_j^{(l)}$  表示第i个顶点  $x_i^{(l)}$  的近邻点,  $\varepsilon^{(l)}$  代表  $x_i^{(l)}$  和  $x_j^{(l)}$  构成的顶点和近邻点的集合,  $h_{\Theta}$  是被可学习的参数  $\Theta$  集合参数化的一系列非线性函数,  $h_{\Theta}^{(l)}$  表示在第1层上的特征学习过程  $\mathbb{R}^{D(1)} \times \mathbb{R}^{D(1)} \rightarrow \mathbb{R}^{D(1+1)}$ 。

$$[0020] \quad x_i^{(l+1)} = \sum_{j:(i,j) \in \varepsilon^{(l)}} h_{\Theta}^{(l)}(x_i^{(l)}, x_j^{(l)}) \quad (5)$$

[0021] (2-2) 使用多窗口和窗口步长的空间金字塔池化结构对学习的点云特征分空间层次地进行多尺度池化;空间金字塔池化结构式如式 (2) 所示:

$$[0022] \quad F(x_1, \dots, x_N) = \underset{x=x_1, \dots, x_N}{mlp} [c\{g(f, s_1), \dots, g(f, s_n)\}] \quad (6)$$

[0023] 其中空间金字塔池化窗口类型包括N、N/4、N/8、N/16,对应点云特征的不同空间层次,其中N为输入的点云个数,f代表通过图卷积网络学习得到的特征信息,  $s_n$  表示第n个池化窗口的尺寸,g代表池化操作,多个不同的池化窗口组合构成了空间金字塔池化结构,c代表将空间金字塔池化得到的多尺度特征进行聚合,经过多层全连接神经网络mlp进行上采样,得到点云中N个点的每个点的特征  $F(x_1, \dots, x_N)$ 。

[0024] (3) 利用3D语义信息、结合基于三维的多视角几何动静点检测算法进行动态物体的检测,剔除动态特征点;

[0025] 本发明使用步骤 (2) 得到的物体语义信息对动态物体进行初步的预筛选,结合基于三维的多视角几何动静点检测方法再次检测是否存在动态特征点,最后去除动态特征点信息,具体步骤如下:

[0026] (3-1) 利用步骤 (2) 得到的环境点云语义信息预筛选可能存在的先验动态物体。

[0027] (3-2) 使用基于三维的多视角几何动静点检测方法检测是否存在动态的特征点,基于三维的多视角几何动静点检测方法引自“DynaSLAM:Tracking,Mapping and Inpainting inDynamic Scenes”。

[0028] 将构建的局部地图中的点投影到当前相机视角下,测量当前帧深度值  $d'$  和投影的深度值  $d$  差值绝对值  $diff$ ,若  $diff$  大于阈值  $TH$ ,说明是有动态物体遮挡了当前场景的部分视野,则判定被遮挡场景前方物体即为动态物体。室内环境中阈值  $TH$  取值范围为  $[0.3, 1]$ ,单位米。

[0029] (3-3) 输出动态物体信息,剔除动态物体中的动态特征点;动态物体信息包括区域位置和对应的特征点信息。

[0030] (4) 求解相机运动,实现视觉里程计,构建并优化局部地图。

[0031] 具体步骤如下:

[0032] (4-1) 对已剔除动态特征点的参考帧和当前帧进行特征匹配,建立帧间几何约束关系求解相机运动,跟踪视觉里程计并生成局部地图,不断选取新的关键帧。若相机跟踪丢失,则进行位姿重定位。

[0033] (4-2) 基于ORB-SLAM2算法,插入新关键帧、剔除冗余地图点、更新创建新的地图点,利用光束平差法(Bundle Adjustment,BA)进行局部地图的相机位姿优化,构建更新局

部地图。

[0034] (5) 通过得到的3D点云语义信息构建目标语义库,具体步骤如下:

[0035] (5-1) 通过语义分割得到的当前点云帧的语义信息,初始化构建包含RGB-D图像、3D包围框、ORB特征点、物体类别和类别概率的语义库。

[0036] (5-2) 对当前点云关键帧进行语义分割,并在语义库中按照物体类别查找是否存在同类别的语义信息。

[0037] (5-3) 若语义库中无与目标物体对应的物体类别信息,说明是初次检测到该物体,则将目标物体添加到语义库中。

[0038] (5-4) 若在超过连续N帧中检测到相同的语义类别,N设为5,即在语义库中找到同类别物体,计算新检测到的物体3D包围框中心与已有同类别所有物体3D包围框中心的欧氏距离,找到距离最小的物体中心坐标,根据位置信息判断是否是同一物体。若是同一个物体信息,说明地图语义库中已存在此目标语义信息,则更新该物体的空间坐标信息,进行新旧物体的地图更新融合。若不是同一个物体,则将此新目标物体加入语义库。

[0039] (6) 根据目标语义库对局部地图进行语义融合,构建八叉树语义地图,具体步骤如下:

[0040] (6-1) 在步骤(4)构建的局部地图基础上,使用贝叶斯法则新融合深度相机的观测数据,进行三维占有格概率动态更新,利用步骤(5)得到的目标语义库不断更新对应位置八叉树节点的语义信息,从而实现八叉树地图的语义化。

[0041] (6-2) 利用光束平差法对构建的八叉树地图进行局部的相机位姿优化和地图配准对齐;然后经过闭环优化模块,用光束平差法对全局地图进行相机位姿和空间点位置的集束调整优化。

[0042] 与现有技术相比,本发明具有以下有益效果:

[0043] 本发明针对ORB-SALM2的定位与建图受环境中动态物体影响的问题,设计了基于3D目标语义分割和三维视觉的多视角几何动静点检测算法相结合的动态物体检测方法,剔除了动态物体掩码中的特征点,然后进行帧间位姿估计和重定位,求解相机运动,减少了动态物体对定位和建图的影响。现有的视觉语义地图构建方法中,3D语义信息获取技术基本停留在通过2D图像的目标检测和语义分割再映射到分割出的点云团上的不足。本发明设计基于动态图卷积和空间金字塔池化的3D点云深度学习网络,在语义地图构建系统中通过直接对3D点云进行语义分割获取目标点云的语义信息的方法,简化了语义的获取流程;构建语义库,设计语义数据关联与融合算法,利用贝叶斯传感器信息融合概率模型进行深度相机的观测数据和地图的融合更新,从而实现了可用于实际导航的八叉树语义地图。

## 附图说明

[0044] 图1本发明流程框图。

[0045] 图2为语义地图构建系统总体框架图。

[0046] 图3为语义地图构建技术细节示意图。

[0047] 图4为点云语义分割网络结构图。

[0048] 图5为基于k-NN的GraphConv示意图。

[0049] 图6为点云空间金字塔池化模型图。

- [0050] 图7为动态物体检测流程图。
- [0051] 图8为基于三维的多视角几合动态点检测示意图。
- [0052] 图9为语义库结构图。
- [0053] 图10为语义库构建流程图。
- [0054] 图11为八叉树结构图。

## 具体实施方式

[0055] 下面结合实施例和附图来详细描述本发明,但本发明并不仅限于此。

[0056] 整个方法系统运行所需的硬件环境为Kinect2.0深度传感器、CPU i7-8700k以及GPU为GTX1080Ti的服务器;软件环境为Ubuntu16.04Linux操作系统、ROS机器人开发环境、ORB-SLAM2开源框架、OpenCV开源视觉库、以及CUDA、cuDNN、Tensorflow等为3D目标检测服务的深度学习环境。另外还有必备的DBoW2视觉词典库、地图显示库Pangolin、图优化库g2o等第三方依赖库。

[0057] 如图1所示,一种基于点云深度学习的室内环境3D语义地图构建方法,主要实施包含以下几个步骤:

- [0058] (1) 使用深度相机获取环境的彩色图像和深度图像;
- [0059] (2) 构建点云深度学习网络获取环境中物体的3D语义信息;
- [0060] (3) 动态物体的检测与动态特征点的剔除;
- [0061] (4) 求解相机运动,实现视觉里程计,构建并优化局部地图;
- [0062] (5) 通过得到的3D语义信息构建目标语义库;
- [0063] (6) 根据目标语义库对局部地图进行语义融合,构建八叉树语义地图。

[0064] 如图2所示为语义地图构建系统总体框架图,选择基于3D点云深度学习的语义分割网络与ORB-SLAM2视觉地图构建系统相结合,设计了一种基于点云深度学习的室内环境3D语义地图构建方法和系统。如图3所示为更详细的3D点云语义分割网络和改进的ORB-SLAM2进行语义融合的语义地图构建技术细节示意图。

[0065] 一种基于点云深度学习的室内环境3D语义地图构建方法,实施步骤(1)具体包括:

[0066] 对Kinect2.0深度相机进行标定,主要分四个部分:彩色图像标定、红外标定、sync帧同步、深度图像标定。然后在Linux环境ROS机器人操作系统下在室内采集色度空间格式为RGB的原始图像和对应的深度图像帧,以及对应的点云帧。

[0067] 步骤(2)构建点云深度学习网络获取环境中物体的3D语义信息,具体实施步骤如下;

[0068] (2-1)将点云深度学习网络PointNet中通过MLP模块对点云进行特征提取的方法改为通过构造动态图卷积提取特征的方法,图的参数随着网络层的不同进行动态更新。

[0069] 点云语义分割网络结构图如图4所示,点云深度学习网络首先采用动态图卷积结构(GraphConv)替换PointNet网络中用多层感知机构造的特征提取结构,GraphConv示意图如图5所示。 $N$ 为采样点的个数, $D$ 为每个点的数据维度, $D=3$ 代表输入三维的点云数据。输入 $N \times D$ 维的点云经过可训练的空间转换网络Spatial T-Net。输入点云经Spatial T-Net网络训练得到的空间转换矩阵进行坐标对齐,然后输入到GraphConv中提取特征。其中 $k$ -NN graph表示用 $k$ 近邻点选取图的范围,mlp( $L_1, L_2, \dots, L_n$ )代表具有 $n$ 个共享权值的MLP层,用来



提取经过k-NN图结构的边缘信息,经过多层图卷积进行特征提取后得到 $N \times k \times L_n$ 维度的特征信息。再将此特征信息经过PSPP进行多尺度池化,使其既包含多尺度的局部区域特征又包含全局特征,用于后续的分割网络。分割网络将经金字塔池化得到的特征与低维度的特征进行融合后再输入全连接层,用来得到每个点的标签分类,实现语义分割。分割网络输出点云中每一点所属m个类别的得分情况。

[0070] 在图5中,首先,对于一个点的个数为n,维度为D的点云,用式(1)表示:

$$[0071] \quad X = \{x_1, \dots, x_n\} \subseteq R^D \quad (1)$$

[0072] 用一个有向图来代表局部点云结构,有向图用式(2)表示,其中 $V = \{1, \dots, n\}$ 代表图结构的顶点, $E \subseteq V \times V$ 表示图结构的边。

$$[0073] \quad G = (V, E) \quad (2)$$

[0074] 通过k-NN图的形式在每一层的点云 $R^D$ 中构建图结构G。对于点 $x_i$ ,找到其k个近邻点用 $x_{j_{i1}}, \dots, x_{j_{ik}}$ 表示,其中 $x_i$ 与邻近点之间的定向边为 $(i, j_{i1}), \dots, (i, j_{ik})$ 。定义边的特征为式(3),其中 $h_{\Theta}$ 是被可学习的参数 $\Theta$ 的集合参数化的一系列非线性函数,用来完成 $R^D \times R^D \rightarrow R^{D'}$ 的特征学习。类比图像中的卷积运算,可以把顶点 $x_i$ 看作做中心像素,只不过此处的 $x_i$ 是三维的,捕获所有的 $x_i$ 就构成了点云全局信息。 $\{x_j: (i, j) \in E\}$ 是环绕在 $x_i$ 周围的点云块, $x_j - x_i$ 代表 $x_i$ 与其邻近点 $x_j$ 构成的局部信息。简言之,输入具有N个点的D维点云通过图卷积后输出具有D'维点云特征的N个点。

$$[0075] \quad e_{ij} = h_{\Theta}(x_i, x_j) \quad (3)$$

[0076] 对于选定顶点 $x_i$ ,以k近邻的方式选择 $x_i$ 的k个邻近点,k取值范围为 $[10, 50]$ ,本发明中当 $k=30$ 时语义分割效果最佳。如图5中仅以 $k=5$ 为例进行示图展示,邻近点在图5中以黄色表示,分别为 $x_{j_{i1}}, x_{j_{i2}}, x_{j_{i3}}, x_{j_{i4}}, x_{j_{i5}}$ 。第二步,通过GraphConv学习得到顶点和邻近点之间的边特征,边特征用 $e_{ij}$ 表示,最后将得到的所有边特征聚合起来。

[0077] 图卷积是动态不固定的,一个点的k近邻点在不同的网络层中是随着特征情况变化的,k近邻点的变化使得每一层的图的结构参数也随着层的不同而进行动态更新,构成了动态的图卷积网络。对于第1层的输出可以表示为式(4):

$$[0078] \quad X^{(l)} = \{x_1^{(l)}, \dots, x_n^{(l)}\} \subseteq R^{D_l} \quad (4)$$

[0079] 原始的点云输入可以表示为 $X^{(0)}$ 。每一层都对应一个不同的图,用式(5)表示:

$$[0080] \quad G = (V^{(l)}, E^{(l)}) \quad (5)$$

$$[0081] \quad x_i^{(l+1)} = \sum_{j:(i,j) \in E^{(l)}} h_{\Theta}^{(l)}(x_i^{(l)}, x_j^{(l)}) \quad (6)$$

[0082] 因此,从第1层的D维度的点云特征中选取的第i个顶点 $x_i^{(l)}$ 与其 $k_1$ 个近邻点经过图卷积学习边缘特征,并经过聚合操作 $\Sigma$ ,得到的第 $(l+1)$ 层的D维度的特征输出 $x_i^{(l+1)}$ 可以用式(6)表示。其中, $x_j^{(l)}$ 表示第i个顶点 $x_i^{(l)}$ 的近邻点, $\epsilon^{(l)}$ 代表 $x_i^{(l)}$ 和 $x_j^{(l)}$ 构成的顶点和近邻点的集合, $h_{\Theta}$ 是被可学习的参数 $\Theta$ 集合参数化的一系列非线性函数, $h_{\Theta}^{(l)}$ 表示在第1层上的特征学习过程 $R^{D^{(l)}} \times R^{D^{(l)}} \rightarrow R^{D^{(l+1)}}$ 。这样,图卷积不仅学会了如何提取局部几何特征,而且学会了如何在点云中对点进行分组。

[0083] 本发明中,局部特征本身没有限定,指的就是局部区域特征,即一帧点云中的局部点特征以及这些特征之间的内在特征联系。

[0084] (2-2)通过多窗口和窗口步长的点云空间金字塔池化(PSPP)结构,如图6所示,对学习的点云特征分空间层次地进行多尺度池化,改进了PointNet网络中对称函数Max pooling模块单一池化窗口的不足。空间金字塔池化结构的数学表达如式(7)所示:

$$[0085] \quad F(x_1, \dots, x_N) = \underset{x=x_1, \dots, x_N}{mlp} [c\{g(f, s_1), \dots, g(f, s_n)\}] \quad (7)$$

[0086] 其中N为输入的点云个数,f代表通过图卷积网络学习得到的特征信息, $s_n$ 表示第n个池化窗口的尺寸,g代表池化操作,多个不同的池化窗口组合构成了空间金字塔池化结构,c代表将空间金字塔池化得到的多尺度特征进行聚合,经过多层全连接神经网络mlp进行上采样,得到点云中N个点的每个点的特征 $F(x_1, \dots, x_N)$ 。

[0087] 具体的,语义分割网络在部件分割数据集ShapeNet上参数设置为网络输入点个数N为2048、构造图卷积结构k近邻点个数为30、空间金字塔池化窗口类型为N,N/4,N/8,N/16的组合学习率为0.001、权重衰减为 $1e-4$ 、动量值0.9、Batch Size为16、Epoch为200;在场景语义分割数据集S3DIS上的参数为输入点个数4096、构造图卷积结构k近邻点个数为30、空间金字塔池化窗口类型为(N,N/4,N/8,N/16)、学习率为0.001、权重衰减为 $1e-5$ 、动量值0.9、Batch Size为24、Epoch为60。

[0088] 然后对ORB-SLAM2进行改进,求解相机运动前,先剔除动态的ORB特征信息,再进行特征匹配,建立帧间几何约束关系求解相机运动,估计帧间位姿与重定位,进行视觉里程计的跟踪和建图。具体实施为步骤(3)和步骤(4)。

[0089] 步骤(3)动态物体的检测与动态特征点的剔除如图7所示,具体包括:

[0090] (3-1)利用步骤(2)得到的环境点云语义信息,预筛选可能存在的先验动态物体,如人等,对于筛选到的动态物体的特征点,在后续的处理中进行剔除,不采用动态物体的特征点求相机的运动。

[0091] (3-2)使用基于三维的多视角几何动静点检测方法检测是否存在动态的特征点,避免了步骤(3-1)中将坐着的人检测为动态物体,同时避免对非先验动态物体的检测遗漏,如手持移动的杯子。

[0092] 如图8基于三维的多视角几何动态点检测示意图:左图所示为无动态遮挡物体下的局部地图示意图,三角形表示参考帧的局部地图点,根据参考帧和当前帧求得的位姿关系,可将局部地图点反投影到当前帧下,没有动态物体出现时,投影点的参考帧投影深度值d和当前帧的测量值d'差接近,但是在右图所示有动态物体遮挡的情况下,测量深度必定小于投影深度,则将对应的点标记为动态点。

[0093] 具体的,将构建的局部地图中的点投影到当前相机视角下,测量当前帧深度值d'和投影的深度值d的差值 $d-d'$ ,若差值大于阈值TH,说明很可能是有动态物体遮挡了当前场景的部分视野,则判定被遮挡场景前方物体即为动态物体。室内环境中,阈值TH取值范围为[0.3,1],单位米。

[0094] (3-3)输出动态物体信息,剔除动态物体对应的动态特征点。根据(3-1)找出先验动态对象,根据步骤(3-2)确定是否存在动态特征点,若存在,则找出这些动态对象的特征点信息,然后剔除运动物体掩码内的特征点,实现动态特征点的剔除。动态物体信息包括区

域位置和对应的特征点信息。

[0095] 步骤(4)具体包括:

[0096] (4-1) 首先利用ORB特征提取算法提取图像特征信息,接着对去除动态特征点的参考帧和当前帧进行特征匹配,建立帧间几何约束关系求解相机运动,然后跟踪视觉里程计生成局部地图,不断选取新的关键帧。若相机跟踪丢失,则进行位姿重定位。

[0097] (4-2) 基于ORB-SLAM2算法,插入新关键帧、剔除冗余地图点、更新创建新的地图点,利用光束平差法进行局部地图的相机位姿优化,构建更新局部地图。

[0098] 步骤(5)通过得到的3D点云语义信息构建目标语义库,语义库结构如图9所示,包含的信息有RGB-D图像、3D包围框、ORB特征点、物体类别和类别概率。具体构建流程如图10所示,步骤如下:

[0099] (5-1) 根据分割得到的语义信息,初始化构建包含RGB-D图像、3D包围框、ORB特征点、物体类别和类别概率的语义库。

[0100] (5-2) 对当前点云关键帧进行语义分割,得到3D语义信息,并在语义库中按照物体类别查找是否存在同类别的语义信息;

[0101] (5-3) 若语义库中无与目标物体对应的物体类别信息,说明是初次检测到该物体,则添加到语义库中;

[0102] (5-4) 若在超过连续帧N中检测到相同的语义类别,设定N为5,即在语义库中找到同类别物体,则计算新检测到的目标3D包围框中心与已有同类别所有物体3D包围框中心的欧氏距离,找到距离最小物体的中心坐标,根据位置信息判断是否是同一物体。若是同一个物体信息,说明地图语义库中已存在此目标语义信息,则更新该物体的空间坐标信息,进行新旧物体的地图更新融合。若不是同一个物体,则将此新物体加入语义库。

[0103] (5-3) 中判断的是没有出现此类别时加入语义库,(5-4) 则是已经出现同类别后的处理,当存在同一语义类别的不同物体时就要更新到语义库中,为下一步更新到地图中做准备。

[0104] 步骤(6)根据目标语义库和局部地图进行语义融合,构建八叉树语义地图。具体包括:

[0105] (6-1) 在步骤(4)构建的局部地图基础上,根据贝叶斯法则更新融合深度相机观测数据,进行三维占有格概率动态更新,利用步骤(5)得到的目标语义库不断更新对应位置八叉树节点的语义信息,实现了八叉树地图的语义化。

[0106] 八叉树结构图如图11所示,八叉树的每个节点都存储着对应空间块的占据信息,也保存了描述一个节点占据程度的空间占有率。当最开始地图是空白的时候,仅需要一个根节点来表示此空白地图,不需要完整的树即可。假设0表示节点是无占据的空白状态,1表示完全占据装状态,用0到1之间的浮点数x作为概率值,表示节点的增大或减小,此浮点型概率值可用一个比特来存储,节省空间。作为概率值,需要对x进行对数变换,避免x超出0到1的范围。假设 $y \in \mathbb{R}$ 为概率对数值,表示一个节点是否被占据。 $x \in [0, 1]$ ,为被占据概率值。其变换和反变换公式分别如下式(8)、(9)所示:

$$[0107] \quad y = \log\left(\frac{x}{1-x}\right) \quad (8)$$

$$[0108] \quad x = \frac{\exp(y)}{\exp(y)+1} \quad (9)$$

[0109] 根据贝叶斯法则更新融合传感器数据的过程,对n节点,观测数据为z,那么从初始时刻到T时刻某节点的概率对数值为 $L(n|z_{1:T})$ ,那么T+1时刻为:

$$[0110] \quad L(n|z_{1:T+1}) = L(n|z_{1:T-1}) + L(n|z_T) \quad (10)$$

[0111] 则第n个节点的占据概率公式为:

$$[0112] \quad P(n|z_{1:T}) = \left[ 1 + \frac{1-P(n|z_T)}{P(n|z_T)} \frac{1-P(n|z_{1:T-1})}{P(n|z_{1:T-1})} \frac{P(n)}{1-P(n)} \right]^{-1} \quad (11)$$

[0113] 节点的先验概率用 $P(n)$ 表示,第n个节点基于T时刻的后验概率用 $P(n|z_T)$ 表示, $z_{1:T}$ 为1到T时间段的观测集合, $P(n|z_{1:T})$ 为第n个节点在1到T时间段观测集合的后验概率。根据上述概率更新公式,可以构建并更新基于八叉树的三维栅格语义地图。

[0114] (6-2) 利用光束平差法(BA)对构建的八叉树地图进行局部地图的位姿优化和配准对齐;然后经过闭环优化模块,用光束平差法对全局地图进行相机位姿和空间点位置的集束调整优化。

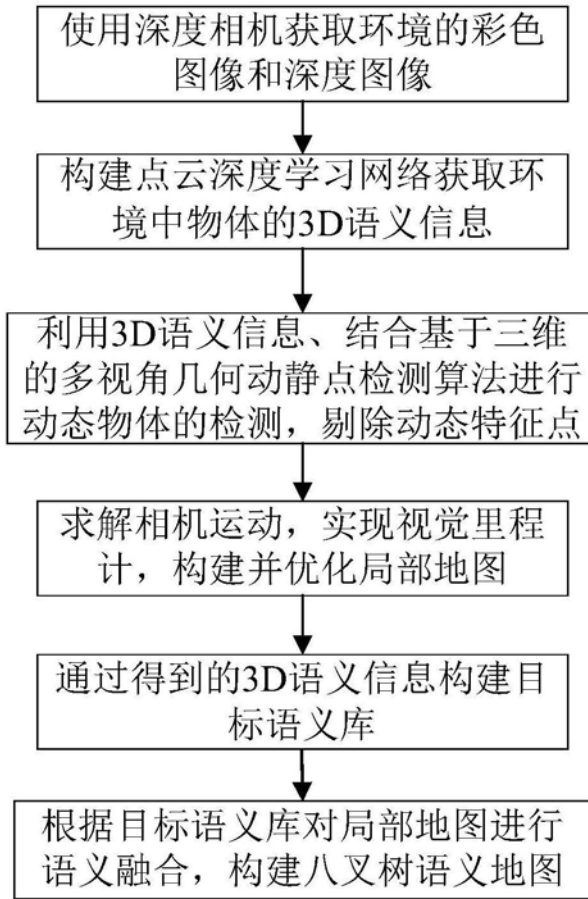


图1

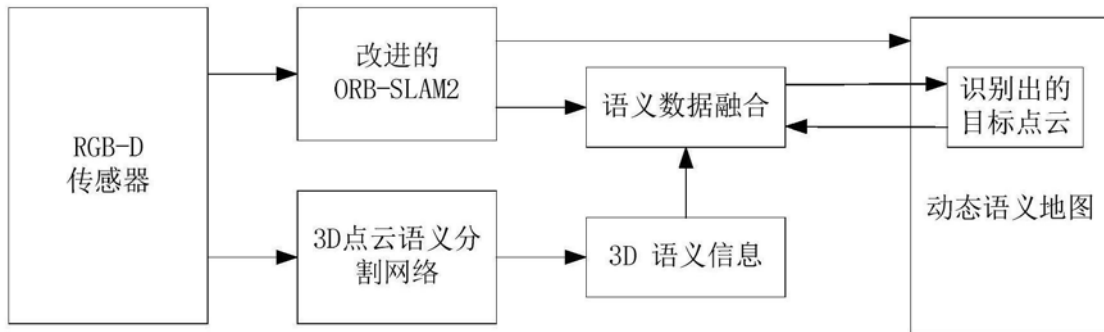


图2

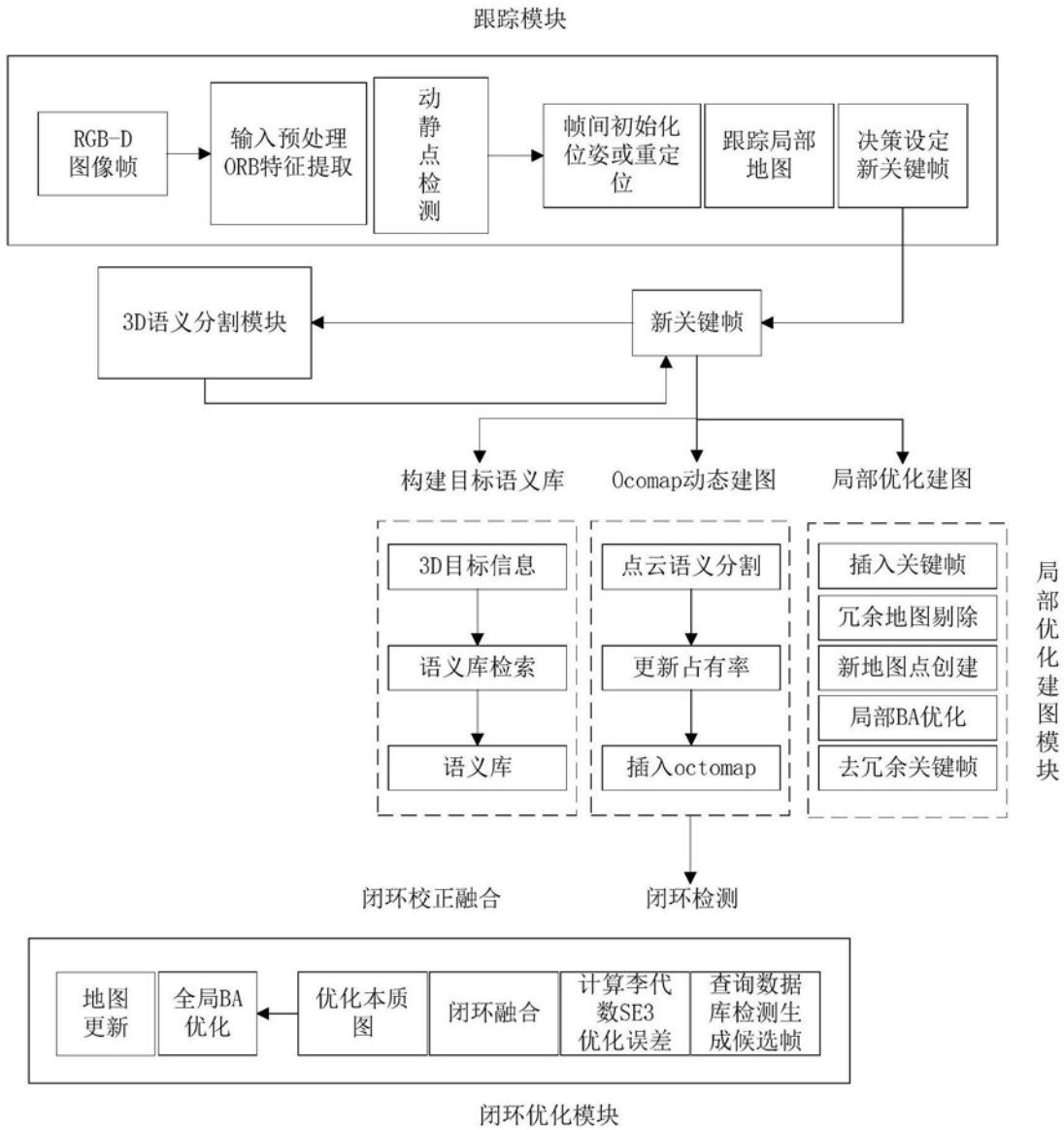


图3

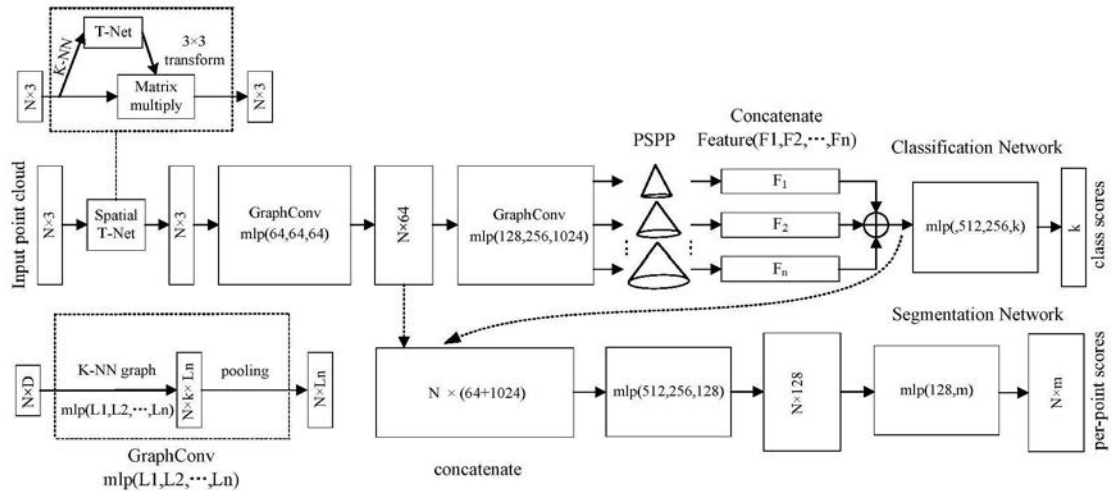


图4

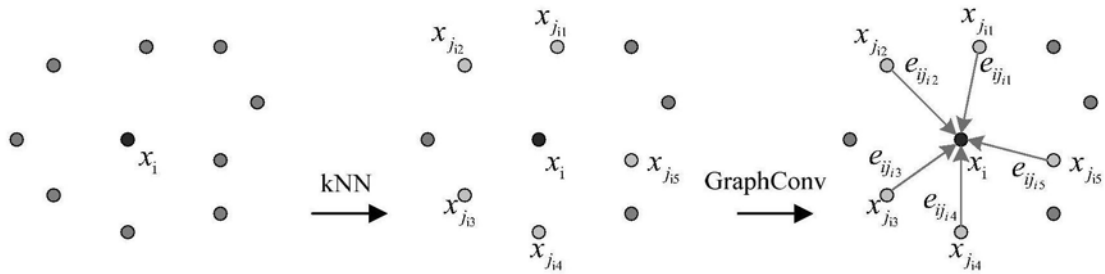


图5

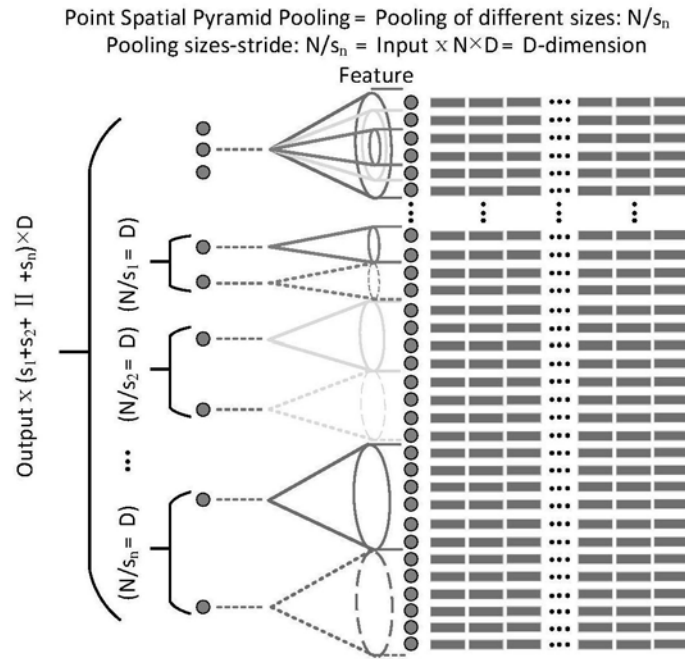


图6



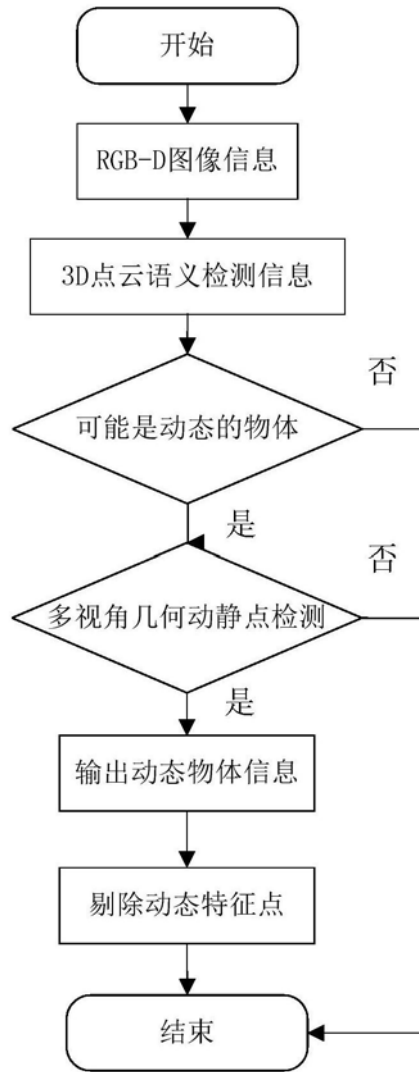


图7

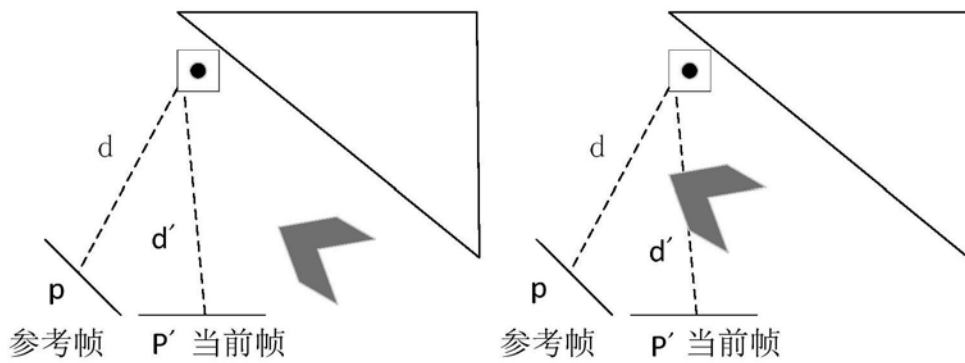


图8

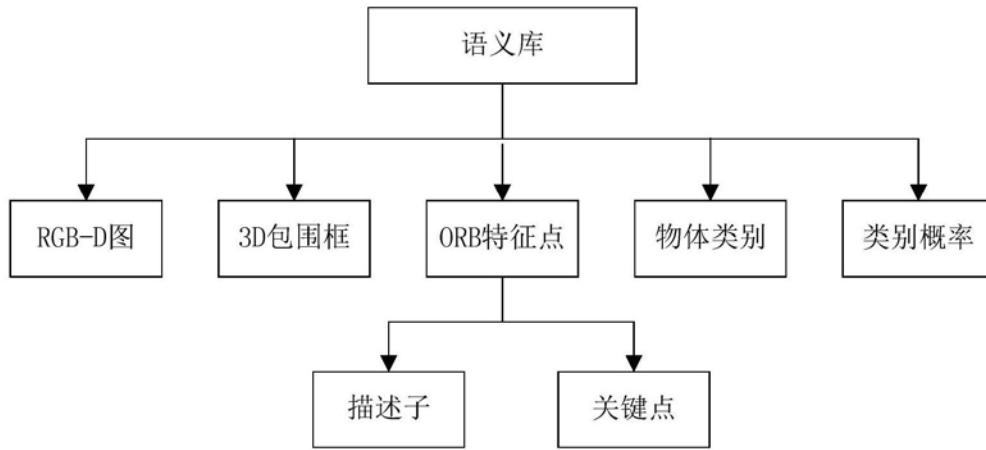


图9

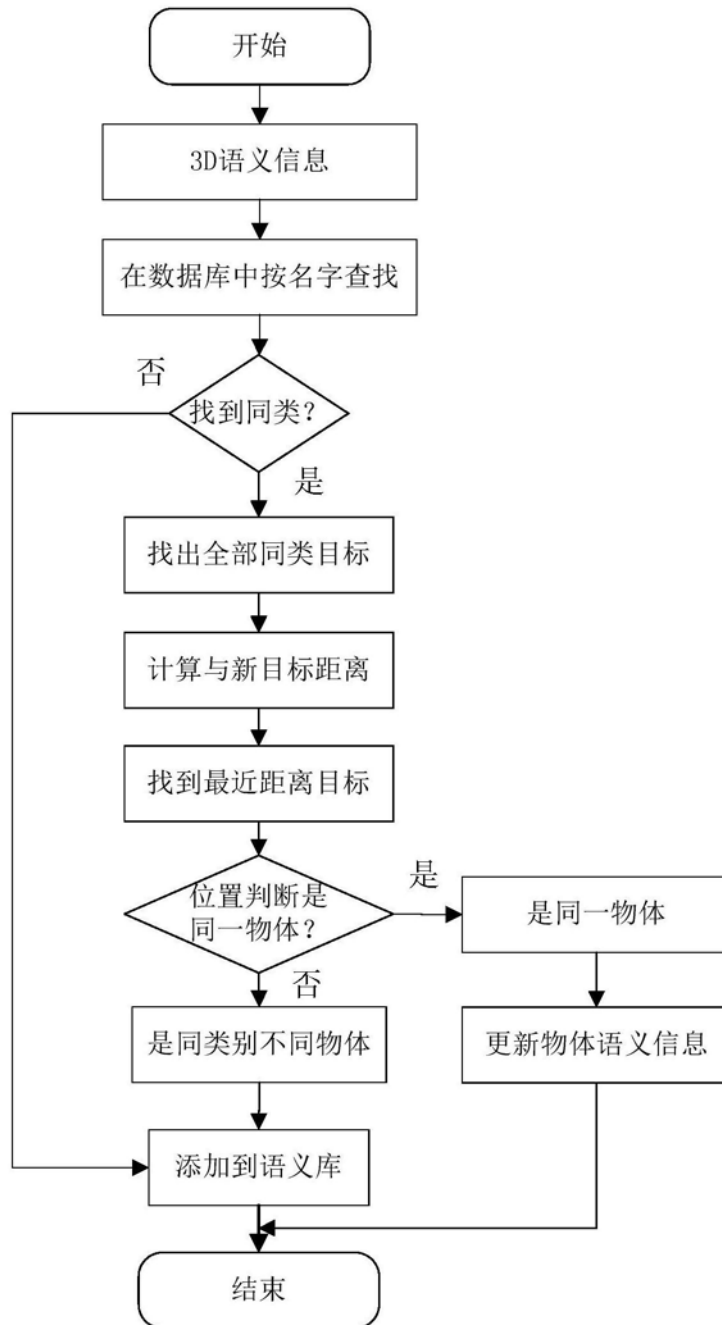


图10

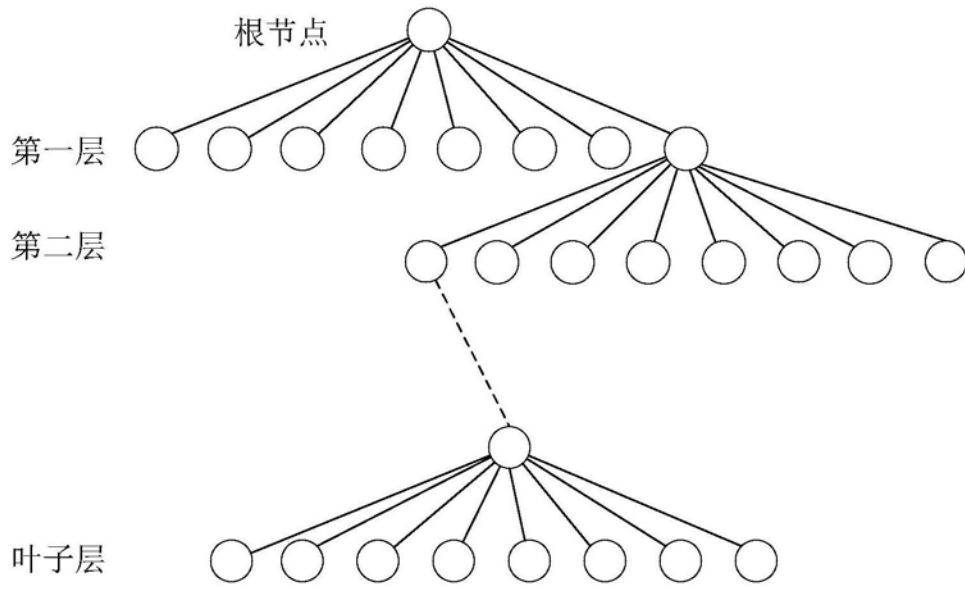


图11