



(12) 发明专利

(10) 授权公告号 CN 115035260 B

(45) 授权公告日 2024. 11. 05

(21) 申请号 202210594240.3

G06V 10/762 (2022.01)

(22) 申请日 2022.05.27

G06N 3/0464 (2023.01)

G06N 3/08 (2023.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 115035260 A

(56) 对比文件

CN 110243370 A, 2019.09.17

CN 112288857 A, 2021.01.29

(43) 申请公布日 2022.09.09

(73) 专利权人 哈尔滨工程大学

审查员 王赞

地址 150001 黑龙江省哈尔滨市南岗区南

通大街145号哈尔滨工程大学科技处

知识产权办公室

(72) 发明人 王立鹏 张佳鹏 王学武 马文龙

张智 张秋雨

(51) Int. Cl.

G06T 17/05 (2011.01)

G06T 7/70 (2017.01)

G06T 7/10 (2017.01)

权利要求书3页 说明书8页 附图7页

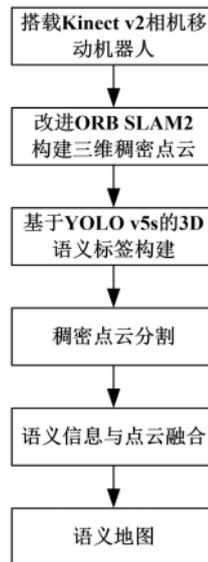
(54) 发明名称

一种室内移动机器人三维语义地图构建方

法

(57) 摘要

本发明公开了一种室内移动机器人三维语义地图构建方法,通过机器人搭载的RGB-D深度相机运行ORB-SLAM2算法,获得机器人关键帧信息与位姿,将关键帧的RGB图与深度图通过反投影形成空间点云;对关键帧的空间点云进行滤波处理并根据位姿进行拼接形成三维稠密点云地图;将关键帧的RGB图利用YOLO V5网络进行目标识别,获得2D语义信息,并根据反投影获得3D语义标签;通过点云分割算法对稠密点云地图进行分割;将获得的语义信息与稠密点云地图的分割结果相融合,构建得到3D语义地图。本发明通过对使用RGB-D深度相机采集的RGB-D信息进行处理,将语义信息与SLAM结果融合,形成信息更为丰富的三维地图。



1. 一种室内移动机器人三维语义地图构建方法,其特征在于,包括以下步骤:

步骤1:通过机器人搭载的RGB-D深度相机运行ORB-SLAM2算法,获得机器人的关键帧信息与对应的位姿,将每一个关键帧对应的RGB图与深度图通过反投影形成单帧空间点云;

步骤2:对关键帧对应的空间点云进行滤波处理,并根据关键帧对应的位姿进行拼接形成三维稠密点云地图;

在ORB-SLAM2系统中的Tracking线程中增加对关键帧的处理,获得关键帧对应的空间点云数据,具体为:对关键帧中的所有像素点进行如下操作,获得相机坐标系下关键帧的点云数据:设图像中某一像素点 P' 坐标为 $[u, v]^T$,相机的内参矩阵为 K ,像素点对应的深度数据为 d ,则将该像素点转换为相机坐标系下的三维点 $P_c(X', Y', Z')$ 为:

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = dK^{-1} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix}$$

得到三维点 P_c 的空间坐标后,再获取像素点 P' 的RGB信息作为 P_c 的RGB信息,对关键帧中的所有像素点进行如上操作,通过PCL库的点云构造函数,得到所述关键帧在相机坐标系下的点云数据;

然后将相机坐标系下关键帧的点云数据转换为世界坐标系下的点云数据:

设 P_c 转换为世界坐标系下的点为 $P_w(X, Y, Z)$, $P_w(X, Y, Z)$ 满足:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = T_{cw}^{-1} \cdot \begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix}$$

其中, T_{cw} 为相机位姿;

步骤3:将关键帧对应的RGB图利用YOLO V5网络进行目标识别,获得关键帧中的2D语义信息,并根据反投影获得3D语义标签;

步骤3.1:建立基于YOLO V5网络模型的数据集

拍摄目标物体图片得到数据集,使用标注程序Labelimg对采集的数据集进行标注,并打上标签,从拍摄图片中分别选取训练集和测试集;从COCO数据集中选取包含目标物体的图片分别作为训练集和测试集;对YOLO V5网络进行数据配置和模型训练参数设置,数据配置包括物体类别数和物体类别名称,利用训练集对YOLO V5网络模型参数进行训练,利用测试集对网络模型进行验证;

步骤3.2:提取2D目标语义和构建3D语义标签

使得YOLO V5网络调用RGB-D深度相机采集RGB图像信息,实时获得RGB图像中的2D识别框作为2D目标语义信息;

在ORB-SLAM2的Tracking线程中添加接口读取实时产生的关键帧,利用训练好的YOLO V5网络模型对每一个新产生的关键帧进行目标检测得到目标物体中心点的像素坐标与边界框的长和宽,然后根据该关键帧对应的RGB图与深度图生成空间点云数据时,将目标物体2D边界框中所有像素点的RGB信息更改为设定的颜色信息,此时根据关键帧生成的单帧点云数据便具有3D语义标签,再将得到的关键帧点云数据结合位姿估计的结果反投影到世界坐标系中完成点云拼接,得到带3D语义标签的全局稠密点云地图,即给点云地图中目标物

体的点云添加颜色信息；

步骤4:通过点云分割算法对稠密点云地图进行分割；

步骤5:将获得的语义信息与稠密点云地图的分割结果相融合,构建得到3D语义地图。

2. 根据权利要求1所述的一种室内移动机器人三维语义地图构建方法,其特征在于:步骤2所述对关键帧对应的空间点云进行滤波处理包括:

使用PCL点云库中的去离群点滤波器去除点云数据中的离群点,对空间点云中每一个点计算其到所有相邻点的平均距离,平均距离在离群距离阈值 d_{pt} 之外的点,认定为离群点并从数据中去除,然后通过体素网格滤波的方法,对点云数据进行降采样。

3. 根据权利要求2所述的一种室内移动机器人三维语义地图构建方法,其特征在于:所述离群距离阈值 d_{pt} 具体为:

设空间点云中所有点的距离服从正态分布 (μ, σ^2) ,判断是否为离群点的阈值参数为 H_{out} ,则离群距离阈值 d_{pt} 为:

$$d_{pt} = \mu + H_{out} \cdot \sigma。$$

4. 根据权利要求1所述的一种室内移动机器人三维语义地图构建方法,其特征在于:步骤2所述根据关键帧对应的位姿进行拼接形成三维稠密点云地图包括:

ORB-SLAM2系统中LoopClosing线程对检测到的闭环中的关键帧位姿进行优化,将闭环优化后的相机位姿发生变化的关键帧提取出来,然后删除世界坐标系下该关键帧对应的点云数据并根据闭环优化后的相机位姿重新生成对应的点云数据,再投影到世界坐标系中,构建出三维稠密点云地图。

5. 根据权利要求1所述的一种室内移动机器人三维语义地图构建方法,其特征在于:步骤4所述通过点云分割算法对稠密点云地图进行分割包括:

步骤4.1:采用超体素聚类分割方法对稠密点云地图进行分割,将点云转换为表面结构,分割结果采用邻接图 $G = \{v, \varepsilon\}$ 表示,其中 $v_i \in v$ 为分割得到的曲面块, ε 表示连接着相邻的曲面块 (v_i, v_j) ,每一个曲面块包含一个质心 c_i 和一个法向量 n_i ;

步骤4.2:在得到曲面块后,使用RANSAC处理曲面块生成候选平面 $PC = \{pc_1, pc_2, \dots, pc_m\}$,然后计算曲面块质心 c_i 到平面 pc_m 的距离 $d(c_i, pc_m)$,设定阈值 δ ,得到所有到平面 pc_m 距离在 δ 内的曲面块,命名为 $\Pi = \{v_i \in V | d(c_i, pc_m) < \delta\}$,定义:

$$D(pc_m) = \begin{cases} 1 - \frac{\Pi}{\eta} & \Pi < \eta \\ \exp(1 - \frac{\eta - \Pi}{\eta}) & \Pi \geq \eta \end{cases}$$

其中, η 表示区分前景物体与背景的约束条件, $D(pc_m)$ 表示点云平面的权重,图分割问题的最小化能量 P^* 为:

$$P^* = \operatorname{argmin} E(P), P \subset PC$$

其中, $E(P)$ 为拟合的能量,满足:

$$E(P) = \sum_{pc_m \in E} D(pc_m)$$

其中 V 与 E 分别为图中顶点与边的集合,对应聚类分割的邻接图 $G = \{v, \varepsilon\}$ 。

6. 根据权利要求1所述的一种室内移动机器人三维语义地图构建方法,其特征在于:步

骤5包括:

根据2D识别框得出目标物体点云的包围盒,其中包围盒的宽度值 b 为2D识别框中全部像素点所对应的最大深度差,假设2D识别框中的全部像素点为集合 $P = \{P_1, P_2, P_3, \dots, P_n\}$,则

$$b = \max(\text{Depth}(P_i) - \text{Depth}(P_j))$$

其中, $P_i, P_j \in P$, $\text{Depth}(P_i)$ 为像素点 P_i 对应的深度值,利用2D像素点反投影到世界坐标系中,将2D识别框的长与宽映射到世界坐标系中得到包围盒的长与高;

然后对点云分割得到的结果进行检测,只有当某个聚类中的所有点云都位于该包围盒中,将聚类的随机颜色属性标签替换为步骤3得到的3D语义标签,获得语义地图。

7. 根据权利要求1至6任意一项所述的一种室内移动机器人三维语义地图构建方法,其特征在于:RGB-D深度相机采用KinectV2相机。

8. 根据权利要求1至6任意一项所述的一种室内移动机器人三维语义地图构建方法,其特征在于:所述YOLOV5网络采用YOLOv5s网络。

一种室内移动机器人三维语义地图构建方法

技术领域

[0001] 本发明属于机器人语义地图构建技术领域,涉及一种机器人地图构建方法,特别是一种室内移动机器人三维语义地图的构建方法。

背景技术

[0002] 随着智能化技术的快速发展,移动机器人产业也迎来了蓬勃发展,先进的移动机器人装载了许多具有特定功能的传感器,在没有人员直接操作的情况下完成各种场景下的指定任务。其中对于应用场景在室内的智能机器人来说,一般会采用机器人自主探索的方式,完成地图构建任务。

[0003] 目前典型的移动机器人语义构建方法主要有基于视觉图像方法和基于点云处理的方法,例如张诚等人在文献《动态环境下的语义SLAM算法》中,利用单目图像序列,开展语义分割、视觉里程计等研究工作,最终构建全局语义地图;再比如张铮等人在文献《变结构的鲁棒语义SLAM算法》中,研究利用点云图像,开展语义SLAM鲁棒性的研究工作。

[0004] 以上现有技术构建移动机器人语义地图方面存在如下问题:一是视觉图像仅能够提供RGB彩色信息,难以提取诸如点云的三维空间信息;二是点云信息不容易获得所代表物体的语义信息,难以将语义与SLAM建图和定位结果相融合。

发明内容

[0005] 针对上述现有技术,本发明要解决的技术问题是提供一种室内移动机器人三维语义地图构建方法,通过对使用RGB-D深度相机采集的RGB-D信息进行处理,将语义信息与SLAM结果融合,形成信息更为丰富的三维地图。

[0006] 为解决上述技术问题,本发明的一种室内移动机器人三维语义地图构建方法,包括以下步骤:

[0007] 步骤1:通过机器人搭载的RGB-D深度相机运行ORB-SLAM2算法,获得机器人的关键帧信息与对应的位姿,将每一个关键帧对应的RGB图与深度图通过反投影形成单帧空间点云;

[0008] 步骤2:对关键帧对应的空间点云进行滤波处理,并根据关键帧对应的位姿进行拼接形成三维稠密点云地图;

[0009] 步骤3:将关键帧对应的RGB图利用YOLO V5网络进行目标识别,获得关键帧中的2D语义信息,并根据反投影获得3D语义标签;

[0010] 步骤4:通过点云分割算法对稠密点云地图进行分割;

[0011] 步骤5:将获得的语义信息与稠密点云地图的分割结果相融合,构建得到3D语义地图。

[0012] 进一步的,步骤1包括:

[0013] 在ORB-SLAM2系统中的Tracking线程中增加对关键帧的处理,获得关键帧对应的空间点云数据,具体为:对关键帧中的所有像素点进行如下操作,获得相机坐标系下关键帧

的点云数据:设图像中某一像素点 P' 坐标为 $[u, v]^T$,相机的内参矩阵为 K ,像素点对应的深度数据为 d ,则将该像素点转换为相机坐标系下的三维点 $P_c(X', Y', Z')$ 为:

$$[0014] \quad \begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = dK^{-1} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix}$$

[0015] 得到三维点 P_c 的空间坐标后,再获取像素点 P' 的RGB信息作为 P_c 的RGB信息,对关键帧中的所有像素点进行如上操作,通过PCL库的点云构造函数,得到所述关键帧在相机坐标系下的点云数据;

[0016] 然后将相机坐标系下关键帧的点云数据转换为世界坐标系下的点云数据:

[0017] 设 P_c 转换为世界坐标系下的点为 $P_w(X, Y, Z)$, $P_w(X, Y, Z)$ 满足:

$$[0018] \quad \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = T_{cw}^{-1} \cdot \begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix}$$

[0019] 其中, T_{cw} 为相机位姿。

[0020] 进一步的,步骤2中对关键帧对应的空间点云进行滤波处理包括:

[0021] 使用PCL点云库中的去离群点滤波器去除点云数据中的离群点,对空间点云中每一个点计算其到所有相邻点的平均距离,平均距离在离群距离阈值 d_{pt} 之外的点,认定为离群点并从数据中去除。然后通过体素网格滤波的方法,对点云数据进行降采样。

[0022] 进一步的,离群距离阈值 d_{pt} 具体为:

[0023] 设空间点云中所有点的距离服从正态分布 (μ, σ^2) ,判断是否为离群点的阈值参数为 H_{out} ,则离群距离阈值 d_{pt} 为:

$$[0024] \quad d_{pt} = \mu + H_{out} \cdot \sigma。$$

[0025] 进一步的,步骤2中根据关键帧对应的位姿进行拼接形成三维稠密点云地图包括:

[0026] ORB-SLAM2系统中LoopClosing线程对检测到的闭环中的关键帧位姿进行优化,将闭环优化后的相机位姿发生变化的关键帧提取出来,然后删除世界坐标系下该关键帧对应的点云数据并根据闭环优化后的相机位姿重新生成对应的点云数据,再投影到世界坐标系中,构建出三维稠密点云地图。

[0027] 进一步的,步骤3包括:

[0028] 步骤3.1:建立基于YOLO V5网络模型的数据集

[0029] 拍摄目标物体图片得到数据集,使用标注程序Labelimg对采集的数据集进行标注,并打上标签,从拍摄图片中分别选取训练集和测试集;从COCO数据集中选取包含目标物体的图片分别作为训练集和测试集;对YOLO V5网络进行数据配置和模型训练参数设置,数据配置包括物体类别数和物体类别名称,利用训练集对YOLO V5网络模型参数进行训练,利用测试集对网络模型进行验证;

[0030] 步骤3.2:提取2D目标语义和构建3D语义标签

[0031] 使得YOLO V5网络调用RGB-D深度相机采集RGB图像信息,实时获得RGB图像中的2D识别框作为2D目标语义信息;

[0032] 在ORB-SLAM2的Tracking线程中添加接口读取实时产生的关键帧,利用训练好的

YOLO V5网络模型对每一个新产生的关键帧进行目标检测得到目标物体中心点的像素坐标与边界框的长和宽,然后根据该关键帧对应的RGB图与深度图生成空间点云数据时,将目标物体2D边界框中所有像素点的RGB信息更改为设定的颜色信息,此时根据关键帧生成的单帧点云数据便具有3D语义标签,再将得到的关键帧点云数据结合位姿估计的结果反投影到世界坐标系中完成点云拼接,得到带3D语义标签的全局稠密点云地图,即给点云地图中目标物体的点云添加颜色信息。

[0033] 进一步的,步骤4中通过点云分割算法对稠密点云地图进行分割包括:

[0034] 步骤4.1:采用超体素聚类分割方法对稠密点云地图进行分割,将点云转换为表面结构,分割结果采用邻接图 $G = \{v, \varepsilon\}$ 表示,其中 $v_i \in v$ 为分割得到的曲面块, ε 表示连接着相邻的曲面块 (v_i, v_j) ,每一个曲面块包含一个质心 c_i 和一个法向量 n_i ;

[0035] 步骤4.2:在得到曲面块后,使用RANSAC处理曲面块生成候选平面 $PC = \{pc_1, pc_2, \dots, pc_m\}$,然后计算曲面块质心 c_i 到平面 pc_m 的距离 $d(c_i, pc_m)$,设定阈值 δ ,得到所有到平面 pc_m 距离在 δ 内的曲面块,命名为 $\Pi = \{v_i \in V \mid d(c_i, pc_m) < \delta\}$,定义:

$$[0036] \quad D(pc_m) = \begin{cases} 1 - \frac{\Pi}{\eta} & \Pi < \eta \\ \exp(1 - \frac{\eta - \Pi}{\eta}) & \Pi \geq \eta \end{cases}$$

[0037] 其中, η 表示区分前景物体与背景的约束条件, $D(pc_m)$ 表示点云平面的权重,图分割问题的最小化能量 P^* 为:

$$[0038] \quad P^* = \arg \min E(P), P \subset PC$$

[0039] 其中, $E(P)$ 为拟合的能量,满足:

$$[0040] \quad E(P) = \sum_{pc_m \in E} D(pc_m)$$

[0041] 其中 V 与 E 分别为图中顶点与边的集合,对应聚类分割的邻接图 $G = \{v, \varepsilon\}$ 。

[0042] 进一步的,步骤5包括:

[0043] 根据2D识别框得出目标物体点云的包围盒,其中包围盒的宽度值 b 为2D识别框中全部像素点所对应的最大深度差,假设2D识别框中的全部像素点为集合 $P = \{P_1, P_2, P_3, \dots, P_n\}$,则

$$[0044] \quad b = \max(\text{Depth}(P_i) - \text{Depth}(P_j))$$

[0045] 其中, $P_i, P_j \in P$, $\text{Depth}(P_i)$ 为像素点 P_i 对应的深度值,利用2D像素点反投影到世界坐标系中,将2D识别框的长与宽映射到世界坐标系中得到包围盒的长与高;

[0046] 然后对点云分割得到的结果进行检测,只有当某个聚类中的所有点云都位于该包围盒中,将聚类的随机颜色属性标签替换为步骤3得到的3D语义标签,获得语义地图。

[0047] 进一步的,RGB-D深度相机采用Kinect V2相机。

[0048] 进一步的,YOLO V5网络采用YOLOv5s网络。

[0049] 本发明的有益效果:本发明通过机器人搭载RGB-D深度相机(例如Kinect相机),获得室内环境的RGB-D信息,基于ORB-SLAM2系统产生的关键帧,通过训练好的YOLO v5s网络模型,完成目标识别获得语义信息后转化为3D语义标签。并根据关键帧对应的RGB图与深度图结合关键帧的位姿信息生成三维稠密点云地图,在此基础上对点云地图进行分割后,将

语义标签与稠密点云地图融合,构建信息丰富的三维语义地图。

附图说明

- [0050] 图1为本发明流程框图;
- [0051] 图2为稠密点云构建流程图;
- [0052] 图3为稠密点云地图;
- [0053] 图4为采集的部分数据集样本;
- [0054] 图5为数据样本的标定;
- [0055] 图6为YOLO v5s实时检测结果;
- [0056] 图7为生成点云标签流程图;
- [0057] 图8为真实场景下点云标签图;
- [0058] 图9为超体素聚类分割结果;
- [0059] 图10为点云分割结果图;
- [0060] 图11为融合后的语义地图。

具体实施方式

[0061] 下面结合说明书附图和实施例对本发明做进一步说明。

[0062] 本发明的流程框图如图1所示,本发明选用Kinect V2相机作为视觉传感器,基于ORB-SLAM 2算法获得相机运动过程中的关键帧,利用关键帧中特征点匹配、后端优化等过程获取相机位姿后,将每一个关键帧的对应的RGB图与深度图通过反投影形成空间点云,经过滤波处理后结合相机的位姿信息进行拼接,形成三维稠密点云地图;同时利用YOLO V5网络模型对关键帧对应的RGB图开展目标识别,获得目标的语义信息,并投影到空间点云实现3D语义标签的构建;然后运用点云分割算法,对稠密点云地图开展分割工作,最后将获得的语义信息与稠密点云地图的分割结果相融合,构建3D语义地图。

[0063] 1、改进ORB-SLAM2构建三维稠密点云

[0064] 本发明通过改进ORB-SLAM2中的部分线程,实现获取稠密点云的目标,具体流程如图2所示,主要步骤如下所示:

[0065] Step 1:二维像素点到三维点云数据反投影

[0066] 对ORB-SLAM2系统中的Tracking线程进行修改,增加了对关键帧的处理,获得关键帧对应的3D点云数据。此过程原理是将环境的二维像素坐标及深度数据通过相机内参和外参反投影到世界坐标系,设图像中某一像素点 P' 坐标为 $[u, v]^T$,相机的内参矩阵为 K ,该像素点对应的深度数据为 d ;则将该像素点转换为相机坐标系下的三维点 $P_c(X', Y', Z')$ 可用下式所示:

$$[0067] \begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = dK^{-1} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \quad (1)$$

[0068] 得到三维点 P_c 的空间坐标后,再获取像素点 P' 的RGB信息作为 P_c 的RGB信息,对关键帧中的所有像素点进行如上操作,通过PCL库的点云构造函数,得到所述关键帧在相机坐标系下的点云数据。得到关键帧对应的点云数据后,将其转换为世界坐标系下的点云,设 P_c 转

换为世界坐标系下的点为 $P_w(X, Y, Z)$, 其过程可以用下式表示, 其中 T_{cw} 为相机位姿:

$$[0069] \quad \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = T_{cw}^{-1} \cdot \begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} \quad (2)$$

[0070] Step2:稠密点云数据滤波

[0071] 本发明使用PCL点云库中的去离群点滤波器去除点云数据中异常的噪声点, 去离群点滤波后得到的点云会比较平滑, 适合后续处理点云聚类。然后通过体素网格滤波的方法, 对点云数据进行降采样。本发明在去离群点滤波时, 对每一个点, 计算它到所有相邻点的平均距离, 得到的结果应服从高斯分布, 其形状是由高斯分布参数 (μ, σ^2) 决定, 平均距离在标准范围之外的点, 认为其为离群点并从数据中去除。

[0072] 设点云中所有点的距离服从正态分布 (μ, σ^2) , 判断是否为离群点的阈值参数为 H_{out} , 则离群距离阈值 d_{pt} 可表示为:

$$[0073] \quad d_{pt} = \mu + H_{out} \cdot \sigma \quad (3)$$

[0074] 即平均距离在区间 $(0, d_{pt})$ 以外的点会被认为是离群点滤除。

[0075] 本发明开展实验, 基于硬件平台(Intel Core i5-10400F CPU, NVIDIA GTX 1650), 将体素网格滤波的分辨率参数调整为0.01, 即体素栅格大小为1cm*1cm*1cm时, 可以获得最佳的运算效果。

[0076] Step3:点云拼接

[0077] 在本发明中, 直接根据step1获得的全局稠密点云数据, 依赖于各个关键帧的相机位姿 T_{cw} , 但ORB-SLAM2系统中LoopClosing线程会根据检测到的闭环对整个闭环中的关键帧位姿进行优化, 此时一些关键帧的位姿估计结果将发生改变。

[0078] 为了避免构建的稠密点云地图出现重影, 本发明将闭环优化后相机位姿发生变化的关键帧提取出来, 然后删除世界坐标系下该关键帧对应的点云数据并根据闭环优化后的位姿重新生成对应的点云数据, 投影到世界坐标系中, 完成回环检测后的点云拼接工作, 最后基于ORB-SLAM2系统构建出的3D稠密点云地图如图3所示。

[0079] 2、基于YOLO v5s网络模型的3D语义标签构建

[0080] 本发明需要对ORB SLAM2系统中Tracking线程生成的关键帧进行目标识别, 然后将得到的2D语义信息映射到空间点云数据中, 完成3D语义标签的构建。具体步骤如下所示:

[0081] Step1:建立基于YOLO v5s网络模型的数据集

[0082] 本专利使用相机拍摄实验室内常见物体, 主要包含电脑设备、书等常见物品。部分采集的图像如图4所示, 使用标注程序Labeling对采集的数据集进行标注, 并打上标签, 如图5所示, 例如标注“显示器”的过程则打上“tv”标签。从拍摄的图片中选取500张作为训练集, 200张作为测试集。此外, 从COCO中选取包含室内场景下物体的图片1500张作为训练集, 300张作为测试集, 为了与COCO数据集一起使用, 将标注文件使用json文件格式存储。在本发明中, 室内环境中常用物体为35种。

[0083] 在训练前需要修改YOLO v5s数据和模型的配置文件, 将两个文件的物体类别数nc改为35, 并在类别列表names里修改为包含的物体类别名称。这里训练集共2000张图片, 测试集共500张图片, 准备好数据集后对YOLO v5s网络模型参数进行训练, 其训练模型的超参数配置如表1所示。

[0084] 表1 YOLO v5s网络模型的参数设置

参数类型	参数值	参数类型	参数值
batch	64	saturation	1.5
subdivisions	32	exposure	1.5
height	640	hue	0.1
width	480	learning_rate	0.0001
channels	3	max_batches	50200
momentum	0.843	policy	steps
decay	0.0005	steps	40000,50000
angle	0	scales	0.1,0.1

[0086] Step2:提取2D目标语义和构建3D语义标签

[0087] YOLO v5s网络训练结束后,使用Kinect V2相机实时采集相应的RGB图像信息,可实现实时检测功能,结果如图6所示,可以获得RGB图像中的2D目标语义信息,能够满足本发明实时构建3D语义标签的要求。

[0088] 将2D语义信息映射到3D语义标签的流程如图7所示,首先是在ORB-SLAM2的Tracking线程中添加接口,基于训练好的YOLO v5s网络,对每一个新产生的关键帧进行目标检测得到目标物体中心点的像素坐标与边界框的长和宽。然后根据该关键帧对应的RGB图与深度图生成空间点云数据时,需将目标物体2D边界框中所有像素点的RGB信息更改为设定的颜色信息,此时根据关键帧生成的单帧点云数据便具有了语义标签。再将得到的关键帧点云数据结合位姿估计的结果反投影到世界坐标系中完成点云拼接,得到带3D语义标签的全局稠密点云地图,即给点云地图中目标物体的点云添加颜色信息。

[0089] 将ORB-SLAM2中生成的关键帧作为输入,获得带有2D边界框和置信度的关键帧图像,再通过反投影得到关键帧中对应的3D语义标签,其结果如图8所示,其中显示器的点云设置为红色,键盘的点云设置为绿色,鼠标的点云设置为蓝色。

[0090] 3、稠密点云分割

[0091] 本发明对构建的稠密点云地图开展点云分割工作,以便获得机器人可理解的场景,同时根据稠密点云地图的分割结果,获得精确的目标点云,便于后续与3D语义标签结合。首先使用基于超体素聚类的分割方法对原始的稠密点云地图进行分割,根据点的相似性,将无规则的点云转换为表面结构,经过超体素分割之后形成的面片都有一个质心与一个法向量,点云分割可以被定义为图分割问题,然后采用图割的方法进行进一步分割,详细步骤如下所示:

[0092] Step1:超体素聚类分割

[0093] 超体素分割是利用点云的空间八叉树结构,通过PCL库中的类Supervoxel Clustering进行聚类,根据分割结果为八叉树结构的体素质心(voxel_centroid_cloud)添加标签,并赋予随机颜色存储在colored_voxel_cloud参数中。

[0094] 其分割原理使用k-均值聚类的区域增长来对点云进行分割,超体素聚类的过程和结晶过程相似,就像溶液过饱和后导致多晶核结晶。超体素聚类的分割方法会在整个空间按

一定规则均匀的分布用来生长的晶核,然后设定好晶核距离和最小晶粒,较小的晶粒被较大的晶粒吸收。在超体素分割过程中,种子点的扩展由特征距离决定,该特征距离考虑了空间、颜色和法向量的特征空间计算,超体素聚类分割中的距离公式如下所示:

$$[0095] \quad D = \sqrt{w_c D_c^2 + \frac{w_s D_s^2}{3R_{seed}^2} + w_n D_n^2} \quad (4)$$

[0096] 上式中: D_c 代表颜色的差异程度, D_n 代表法线方向的差异程度, D_s 表示点云在距离上的差异程度。公式中 w_c 、 w_s 、 w_n 分别表示各个变量的权重。通过对晶核周围的搜索,寻找D值最小的体素即为最相似的,以此来代表下一个进行生长的晶核。

[0097] 本发明选择Kinect V2相机作为传感器,设置合适的参数: $voxel_resolution=0.008$ (体素大小,空间八叉树的分辨率), $seed_resolution=0.5$ (种子分辨率)。根据实验室环境调整颜色、空间距离和法向量所占的权重,获得的超体素聚类分割结果如图9所示。

[0098] Step2:基于几何信息的点云分割

[0099] 上一步获得的超体素分割结果采用邻接图 $G = \{v, \varepsilon\}$ 表示,其中 $v_i \in v$ 为分割得到的曲面块, ε 表示连接着相邻的曲面块(v_i, v_j)。经过超体素分割处理之后,每一个曲面块都有一个质心 c_i ,和一个法向量 n_i ,此时场景分割可视为图分割问题。

[0100] 假设一个点云地图中有K个实际平面 $\{s_1, s_2, \dots, s_k\}$,并且这些平面已经被分割成曲面块。这里定义一个变量 $\{b_i\}_1^N, b_i \in [0, K], b_i = K$ 表示曲面块属于平面 s_k, N 代表了实际平面 s_k 被分割成了N个曲面块,如图9中桌面分成了两个聚类。若能根据点云地图提取其中所有对象的平面并将曲面块分配给它们,便可得到较为精确的点云分割结果。

[0101] 具体操作在得到曲面块后,使用RANSAC处理曲面块以生成候选平面 $PC = \{pc_1, pc_2, \dots, pc_m\}$,计算 $d(c_i, pc_m)$,即曲面块质心 c_i 到平面 pc_m 的距离。增加一个阈值 δ ,可以得到所有到平面 pc_m 距离在 δ 内的曲面块,命名为 $\Pi = \{v_i \in V \mid d(c_i, pc_m) < \delta\}$ 。定义下式:

$$[0102] \quad D(pc_m) = \begin{cases} 1 - \frac{\Pi}{\eta} & \Pi < \eta \\ \exp(1 - \frac{\eta - \Pi}{\eta}) & \Pi \geq \eta \end{cases} \quad (5)$$

[0103] 上式中: η 表示区分前景物体与背景的约束条件, $D(pc_m)$ 表示点云平面的权重。在实验中,设置 $\eta=40$ 和 $\delta=4$ (cm)。图分割问题的最小化能量 P^* 可用下式表示:

$$[0104] \quad P^* = \arg \min E(P), P \subset PC \quad (6)$$

[0105] 上式中 $E(P)$ 为拟合的能量,具体如下式所示:

$$[0106] \quad E(P) = \sum_{pc_m \in E} D(pc_m) \quad (7)$$

[0107] 其中 V 与 E 分别为图中顶点与边的集合,对应聚类分割的邻接图 $G = \{v, \varepsilon\}$ 。

[0108] 基于点云最小割理论切断满足能量最小化时的边,也即能够根据约束条件合并相邻的曲面块便可得到点云的分割结果,此时的分割目的是为了将聚类结果中属于同一物体的聚类合并,分割结果如图10所示,其中,目标物体如显示器、书本、杯子、瓶子等分割效果很好,背景物体如桌子、置物柜、地面等平面的合并效果也很好,满足本发明的要求。

[0109] 4、融合语义信息,构建语义地图

[0110] 本发明在完成稠密点云分割后,将分割结果与3D语义标签相结合,通过语义信息来优化分割结果,可将点云分割中未能提取出的目标分割出来;同时,可为分割出的各个目标添加语义信息,构建更为精确的语义地图。点云分割之后,每一个聚类均具有一个属性标签,用特定目标的语义标签替换掉聚类的属性标签之后,点云中的特定对象便具有了语义信息,从而构建出语义地图。具体实施步骤如下所示:

[0111] Step1:语义标签的存储

[0112] 本发明在对稠密点云地图进行分割后,不同颜色的点云代表着每个超体素中的点云集合和空间采样区域,此颜色便为该聚类的属性标签。但此时的属性标签仅仅是区分出各个聚类结果,其本身是一个没有任何意义的随机标签,此标签亦为最终点云分割结果中各个聚类的属性标签。

[0113] 在第二部分中,本发明通过YOLO v5s对ORB-SLAM2中产生的关键帧进行目标识别,并根据识别结果获得的3D语义标签,为点云分割的结果添加带语义信息的颜色标签,便完成了对物体语义标签的存储。

[0114] 不过此时是将2D目标识别边界框中的所有像素点结合Kinect相机的深度数据直接映射到3D形成的语义标签。在实验室的复杂环境中,目标识别的边界框不够精准,很容易导致目标物体的周围点云添加错误的语义信息,如图8所示,存在将语义标签错误的添加到了目标物体的周围点云中的情况。

[0115] Step2:语义标签与点云包围盒的融合

[0116] 为了获得更为精确的语义地图,本发明根据2D识别框得出目标物体点云的包围盒,其中包围盒的宽度值 b 为2D识别框中全部像素点所对应的最大深度差。假设2D识别框中的全部像素点为集合 $P = \{P_1, P_2, P_3, \dots, P_n\}$,则

[0117]
$$b = \max(\text{Depth}(P_i) - \text{Depth}(P_j)) \quad (8)$$

[0118] 其中 $P_i, P_j \in P$, $\text{Depth}(P_i)$ 为像素点 P_i 对应的深度值。包围盒的长与高由2D识别框的长与宽映射得到,原理为2D像素点反投影到世界坐标系中,在此不再赘述。

[0119] 接下来,本发明将对点云分割得到的结果进行检测,只有当某个聚类中的所有点云都位于该包围盒中,才会被添加特定的语义信息,即为其添加特定的颜色属性,由此获得更为精确的语义地图。如图11与图10所示,通过语义标签获得的包围盒与点云分割结果的融合,物体TV能够分割的更为完整,实际物体边缘点云的聚类也成功添加了正确的语义信息;与图8直接通过2D目标识别获得的边界框为点云添加语义信息相比,本方法也能够有效避免将语义信息错误的添加到目标周围的点云上。同时,结合语义信息也能够改善点云分割效果,如图11中,添加的语义信息改变颜色属性后的再聚类,能够分割出桌面上的键盘、鼠标等小物体。按照如上两个步骤,便可以形成如图11所示的三维语义地图。

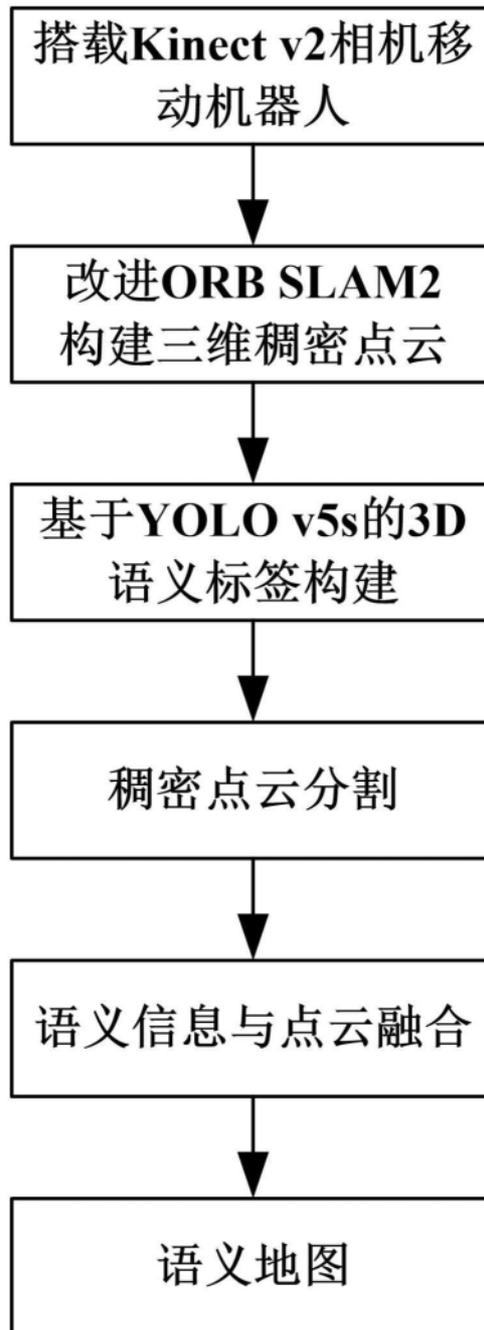


图1

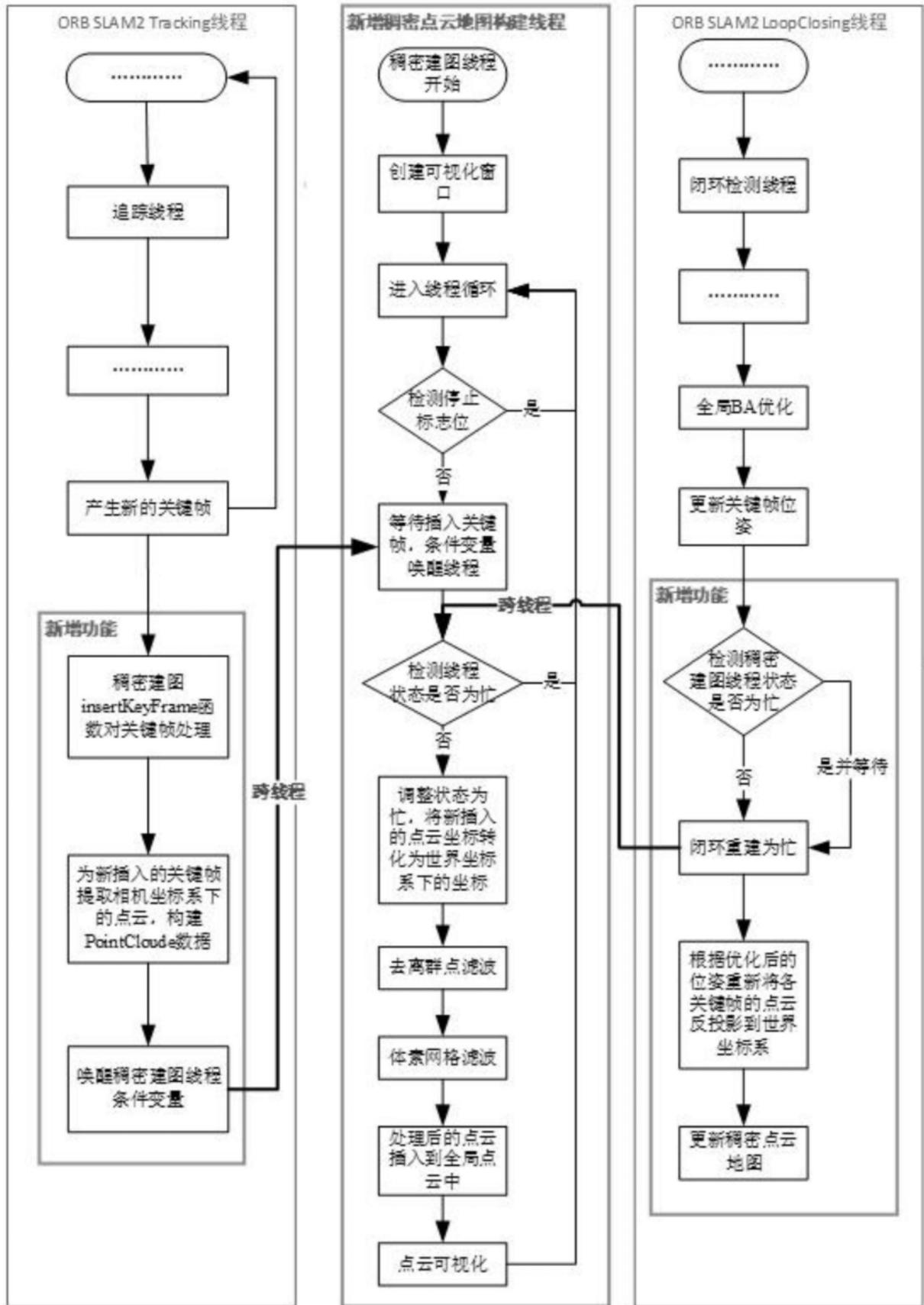


图2

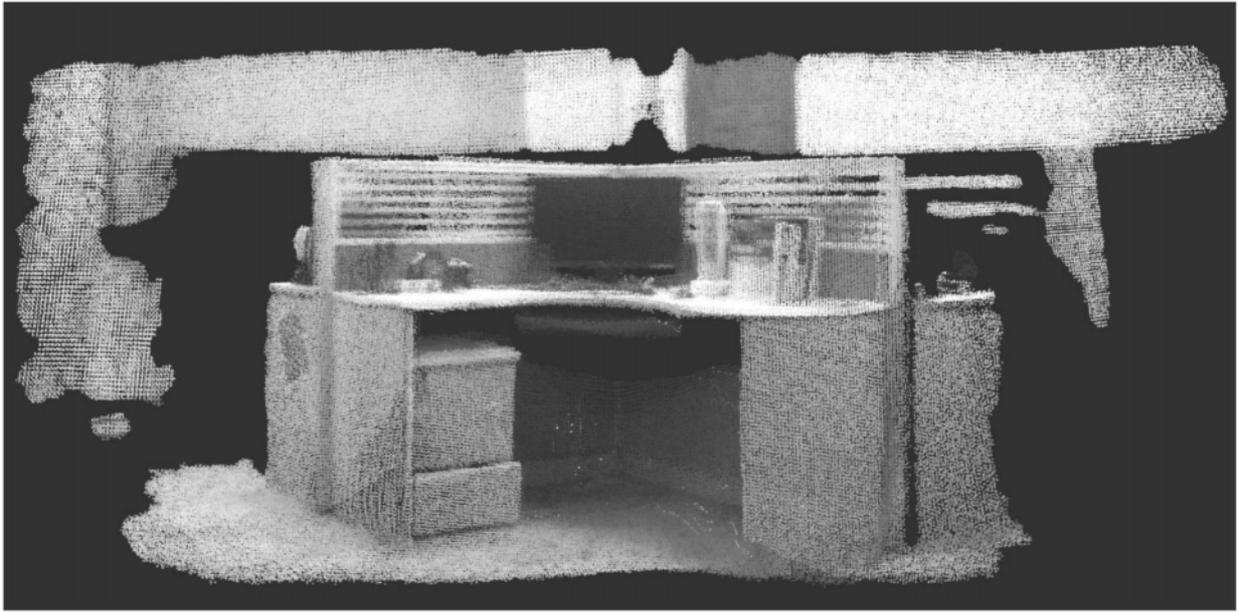


图3

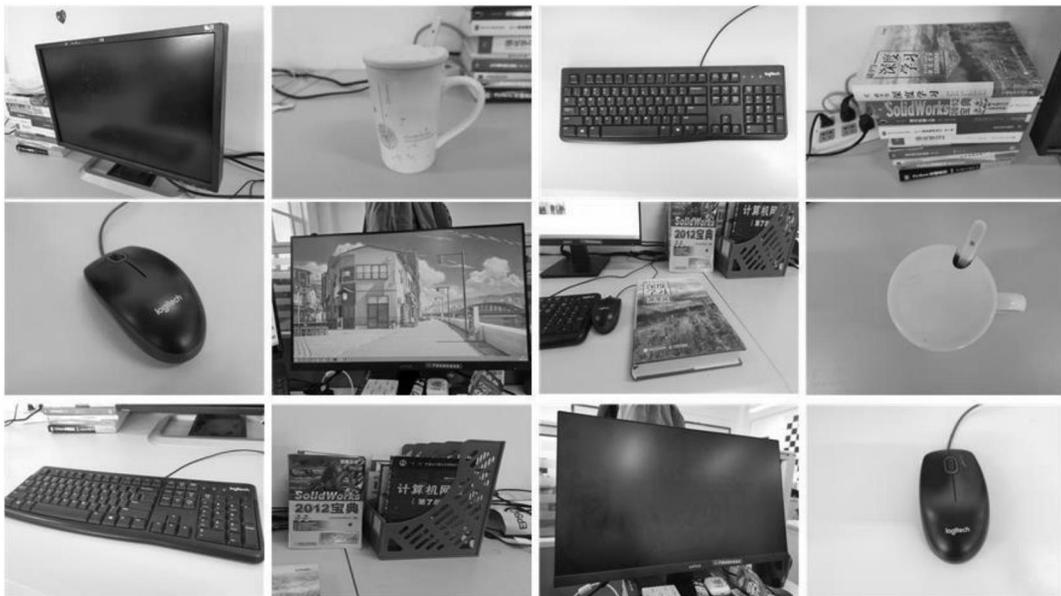


图4



图5

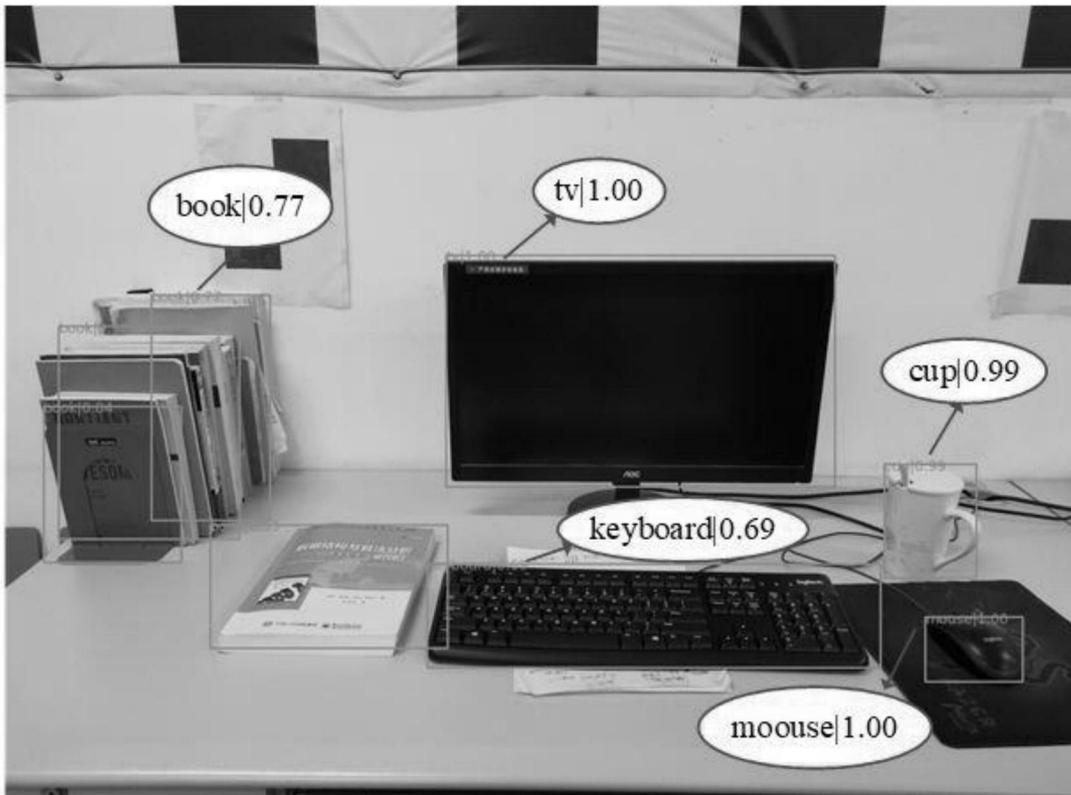


图6

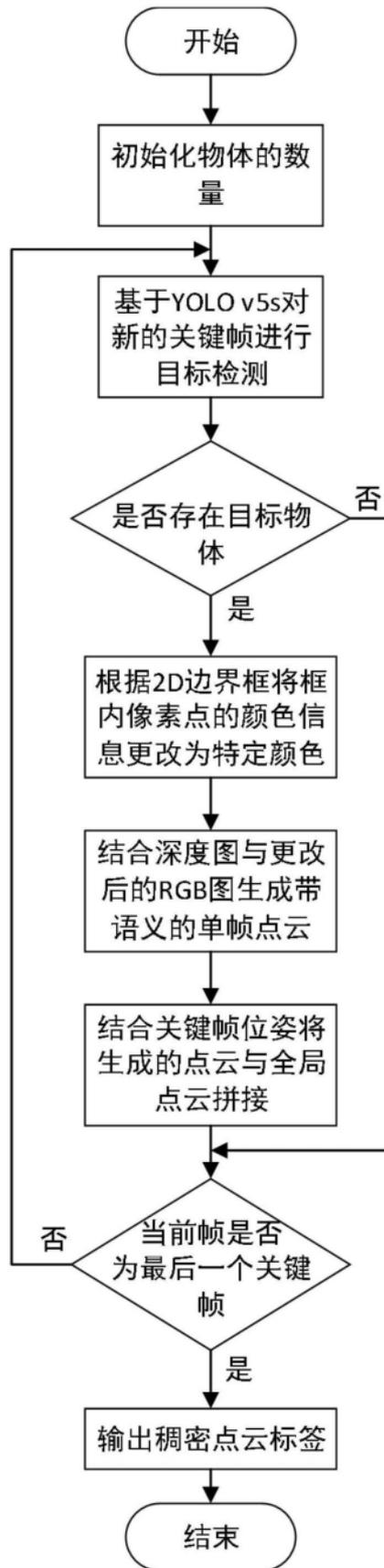


图7

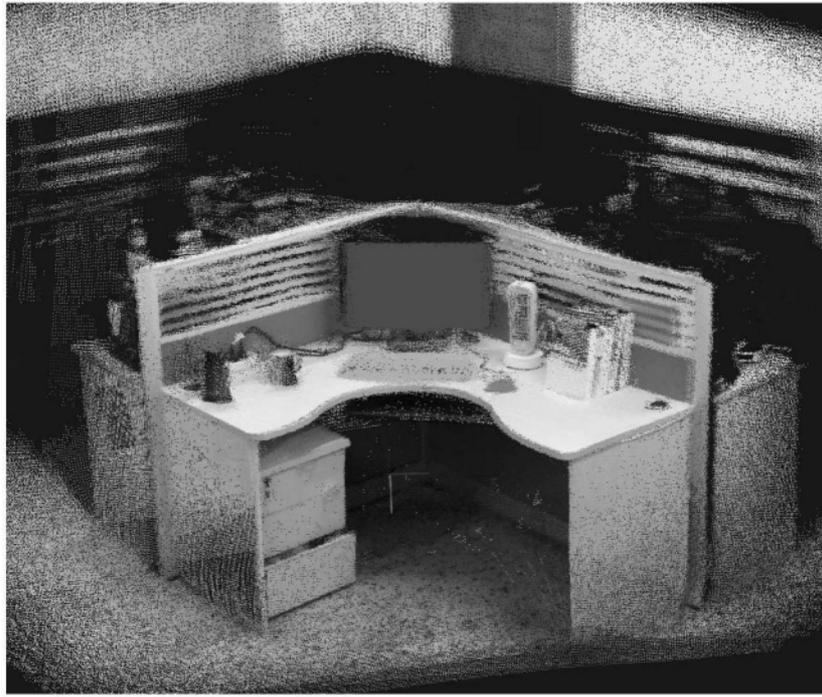


图8

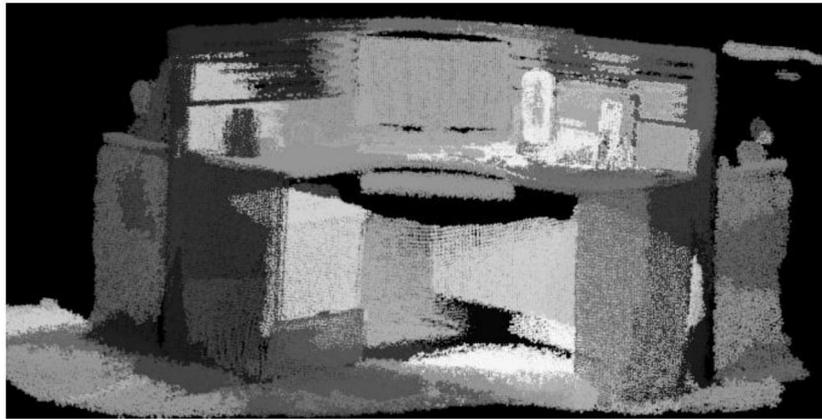


图9

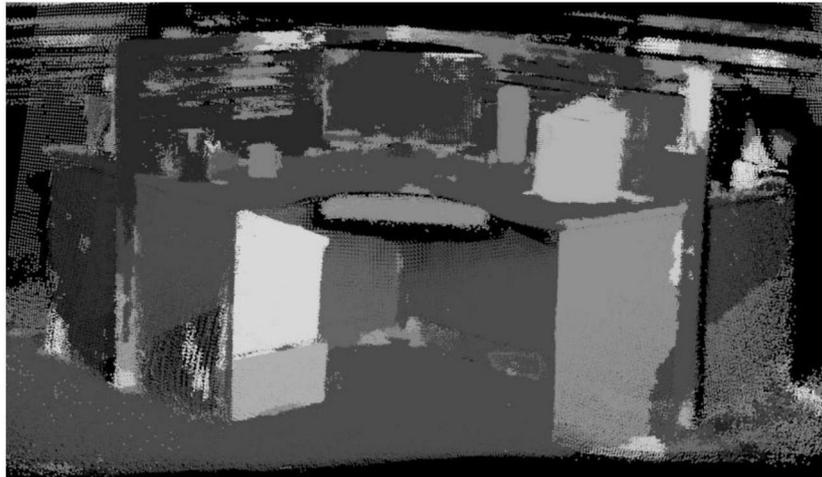


图10

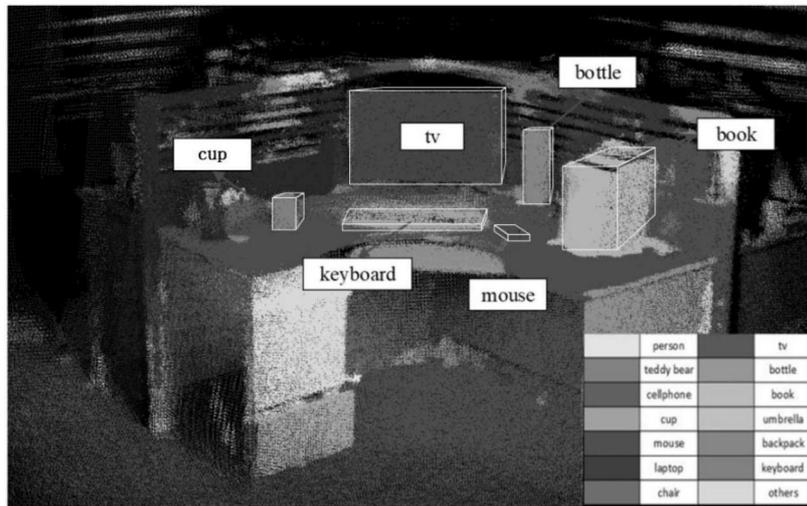


图11