(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 109559320 A
(43)申请公布日 2019.04.02

(21)申请号 201811388678.6

(22)申请日 2018.11.21

(66)本国优先权数据
201811088531.5 2018.09.18 CN

(71)申请人 华东理工大学
地址 200237 上海市徐汇区梅陇路130号

(72)发明人 朱煜 黄俊健 陈旭东 郑兵兵 倪光耀

(74)专利代理机构 上海智信专利代理有限公司 31002
代理人 王洁 郑暄

(51)Int.Cl.
G06T 7/11(2017.01)
G06K 9/62(2006.01)
G06N 3/04(2006.01)

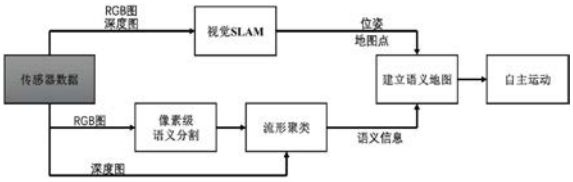
权利要求书4页 说明书11页 附图2页

(54)发明名称

基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法及系统

(57)摘要

本发明涉及一种基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法，包括(1)嵌入式开发处理器通过RGB-D摄像头获取当前环境的彩色信息与深度信息；(2)通过采集的图像得到特征点匹配对，并进行位姿估计，且获得场景空间点云数据；(3)利用深度学习对图像进行像素级语义分割，通过图像坐标系和世界坐标系映射，并使得空间点具有语义标注信息；(4)通过流形聚类消除优化语义分割所带来的误差；(5)进行语义建图，对空间点云进行拼接，得到由密集离散的点组成的点云语义地图。本发明还涉及一种基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的系统。采用了该方法及系统，空间网络地图具有更高级的语义信息，更符合在实时建图过程中的使用需求。



CN 109559320 A

1.一种基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法,其特征在于,所述的方法包括以下步骤:

(1)嵌入式开发处理器通过RGB-D摄像头获取当前环境的彩色信息与深度信息;

(2)通过采集的图像得到特征点匹配对,并进行位姿估计,且获得场景空间点云数据;

(3)利用深度学习对图像进行像素级语义分割,通过图像坐标系和世界坐标系映射,并使得空间点具有语义标注信息;

(4)通过流形聚类消除优化语义分割所带来的误差;

(5)进行语义建图,对空间点云进行拼接,得到由密集离散的点组成的点云语义地图。

2.根据权利要求1所述的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法,其特征在于,所述的步骤(1)中的嵌入式处理器包括NVIDIA JETSON TX2系统。

3.根据权利要求1所述的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法,其特征在于,所述的步骤(2)包括以下步骤:

(2.1)通过视觉SLAM技术提取图像特征点,进行特征匹配得到特征点匹配对;

(2.2)通过3D点对求解相机当前位姿;

(2.3)通过图优化Bundle Adjustment的方法进行更精确的位姿估计;

(2.4)通过回环检测消除帧间的累计误差,并获得场景空间点云数据。

4.根据权利要求1所述的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法,其特征在于,所述的步骤(3)中的对图像进行像素级语义分割具体包括以下步骤:

(3.1)通过基于改进空洞卷积的GoogLeNet的特征提取层;

(3.2)通过基于改进空洞卷积的GoogLeNet的多尺度提取层;

(3.3)根据提取结果对图像进行分类。

5.根据权利要求4所述的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法,其特征在于,所述的步骤(3.1)还包括特征提取层的设计过程,具体包括以下步骤:

(3.1.1)将GoogLeNet网络结构中Inception(3b)之后的最大池化层步长修改为1;

(3.1.2)将GoogLeNet网络结构中Inception(4a)、Inception(4b)、Inception(4c)、Inception(4d)、Inception(4e)部分使用空洞卷积代替,并设置空洞卷积为5×5且dilation为2的Pool;

(3.1.3)将GoogLeNet网络结构中Inception(4e)之后的最大池化层步长修改为1。

6.根据权利要求4所述的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法,其特征在于,所述的步骤(3.2)还包括多尺度提取层的设计过程,具体包括以下步骤:

(3.2.1)基于空间金字塔池化进行多尺度处理;

(3.2.2)通过1×1卷积和不同采样率的空洞卷积提取不同尺度的特征图像;

(3.2.3)融合图像池化特征到模块中,将所述的特征图像经过1×1的卷积融合得到特征,并放入Softmax层进行像素点语义分类。

7.根据权利要求1所述的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法,其特征在于,所述的步骤(4)具体包括以下步骤:

(4.1)计算空间点的切平面法向量;

(4.2)搜索未被分配类别的点x_i,判断是否所有点均已聚类,如果是,则继续步骤(4.5);否则,x_i类别为c=c+1,并创建空队列q;

(4.3)计算空间点x_i的切平面法向量v_i和距其小于0.01范围内所有点x_j的法向量v_j的夹角α_{ij},判断是否存在$\alpha_{ij}<\sigma$或者$\alpha_{ij}>175°$,如果是,则x_j和x_i归为一类,x_j类别为c,并将满足条件的x_j压入队列q中;否则,继续步骤(4.4);

(4.4)判断队列q是否非空,如果是,则令$x_i=q_1$,继续步骤(4.3);否则继续步骤(4.1);

(4.5)提取簇内点数最多的前k类点,剩下的点按照就近原则归类。

8.根据权利要求1所述的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法,其特征在于,所述的步骤(4.1)中的计算空间点的切平面法向量,具体为:

根据以下公式计算空间点的切平面法向量:

$$w = argmin_w \frac{1}{2} aw^T w = argmin_w \frac{1}{2} a,$$

其中,$w\in R^{3\times1}$为该平面的单位法向量,a为特征值。

9.根据权利要求1所述的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法,其特征在于,所述的步骤(5)包括以下步骤:

(5.1)根据RGB-D相机的精度特性,去除深度值太大或无效的点云;

(5.2)通过统计滤波器方法去除孤立的空间点,计算每个空间点与它最近N个空间点的距离均值,去除距离均值过大的空间点;

(5.3)通过空间网格原理,将所有空间点云填充进空间网格,使得每个空间网格只保留一个空间点。

10.一种基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的系统,其特征在于,所述的系统包括:

嵌入式开发处理器,用于构建视觉SLAM语义地图;

RGB-D相机,与所述的嵌入式开发处理器相连接,用于采集彩色数据和深度数据;

建图程序,所述的建图程序在运行时根据深度学习与视觉SLAM,通过嵌入式开发处理器和RGB-D相机实现视觉SLAM语义建图,具体进行以下步骤处理:

(1)嵌入式开发处理器通过RGB-D摄像头获取当前环境的彩色信息与深度信息;

(2)通过采集的图像得到特征点匹配对,并进行位姿估计,且获得场景空间点云数据;

(3)利用深度学习对图像进行像素级语义分割,通过图像坐标系和世界坐标系映射,并使得空间点具有语义标注信息;

(4)通过流形聚类消除优化语义分割所带来的误差;

(5)进行语义建图,对空间点云进行拼接,得到由密集离散的点组成的点云语义地图。

11.根据权利要求10所述的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的系统,其特征在于,所述的步骤(1)中的嵌入式处理器包括NVIDIA JETSON TX2系统。

12.根据权利要求10所述的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的系统,其特征在于,所述的步骤(2)包括以下步骤:

(2.1)通过视觉SLAM技术提取图像特征点,进行特征匹配得到特征点匹配对;

(2.2)通过3D点对求解相机当前位姿;

(2.3)通过图优化Bundle Adjustment的方法进行更精确的位姿估计;

(2.4)通过回环检测消除帧间的累计误差,并获得场景空间点云数据。

13.根据权利要求10所述的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的

系统，其特征在于，所述的步骤(3)中的对图像进行像素级语义分割具体包括以下步骤：

(3.1)通过基于改进空洞卷积的GoogLeNet的特征提取层；

(3.2)通过基于改进空洞卷积的GoogLeNet的多尺度提取层；

(3.3)根据提取结果对图像进行分类。

14.根据权利要求13所述的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的系统，其特征在于，所述的步骤(3.1)还包括特征提取层的设计过程，具体包括以下步骤：

(3.1.1)将GoogLeNet网络结构中Inception(3b)之后的最大池化层步长修改为1；

(3.1.2)将GoogLeNet网络结构中Inception(4a)、Inception(4b)、Inception(4c)、Inception(4d)、Inception(4e)部分使用空洞卷积代替，并设置空洞卷积为5×5且dilation为2的Pool；

(3.1.3)将GoogLeNet网络结构中Inception(4e)之后的最大池化层步长修改为1。

15.根据权利要求13所述的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的系统，其特征在于，所述的步骤(3.2)还包括多尺度提取层的设计过程，具体包括以下步骤：

(3.2.1)基于空间金字塔池化进行多尺度处理；

(3.2.2)通过1×1卷积和不同采样率的空洞卷积提取不同尺度的特征图像；

(3.2.3)融合图像池化特征到模块中，将所述的特征图像经过1×1的卷积融合得到特征，并放入Softmax层进行像素点语义分类。

16.根据权利要求10所述的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的系统，其特征在于，所述的步骤(4)具体包括以下步骤：

(4.1)计算空间点的切平面法向量；

(4.2)搜索未被分配类别的点x_i，判断是否所有点均已聚类，如果是，则继续步骤(4.5)；否则，x_i类别为c＝c+1，并创建空队列q；

(4.3)计算空间点x_i的切平面法向量v_i和距其小于0.01范围内所有点x_j的法向量v_j的夹角α_{ij}，判断是否存在$\alpha_{ij}<\sigma$或者$\alpha_{ij}>175°$，如果是，则x_j和x_i归为一类，x_j类别为c，并将满足条件的x_j压入队列q中；否则，继续步骤(4.4)；

(4.4)判断队列q是否非空，如果是，则令$x_i=q_1$，继续步骤(4.3)；否则继续步骤(4.1)；

(4.5)提取簇内点数最多的前k类点，剩下的点按照就近原则归类。

17.根据权利要求10所述的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的系统，其特征在于，所述的步骤(4.1)中的计算空间点的切平面法向量，具体为：

根据以下公式计算空间点的切平面法向量：

$$w = argmin_w \frac{1}{2} aw^T w = argmin_w \frac{1}{2} a,$$

其中，$w \in R^{3\times1}$为该平面的单位法向量，a为特征值。

18.根据权利要求10所述的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的系统，其特征在于，所述的步骤(5)包括以下步骤：

(5.1)根据RGB-D相机的精度特性，去除深度值太大或无效的点云；

(5.2)通过统计滤波器方法去除孤立的空间点，计算每个空间点与它最近N个空间点的距离均值，去除距离均值过大的空间点；

(5.3)通过空间网格原理，将所有空间点云填充进空间网格，使得每个空间网格只保留

一个空间点。

基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法及系统

技术领域

[0001] 本发明涉及无人系统实时定位与建图领域，尤其涉及图像处理的语义分割领域，具体是指一种基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法及系统。

背景技术

[0002] 近年来无人系统发展迅速，自动驾驶、机器人和无人机都是典型的无人系统。视觉SLAM(Simultaneous Localization and Mapping，即时定位与建图)系统已被广泛的应用于无人系统的定位与路径规划中，如由Mur-Artal等于2015年提出的ORB-SLAM(Mur-Artal R,Montiel J M M,Tardós J D.ORB-SLAM:A Versatile and Accurate Monocular SLAM System[J].IEEE Transactions on Robotics,2015,31(5):1147-116)。视觉SLAM系统中所建立的空间网络地图仅仅包含低级信息，比如色彩信息和距离信息，这样不利于机器人对当前场景的理解，所以我们在视觉SLAM系统的构建过程中引入基于深度学习的语义分割网络，实现机器人对当前场景的语义及场景理解。

[0003] 语义分割的目的是用于场景理解，实现了各类目标之间的精确分割，可以用于自动驾驶或者机器人来帮助识别目标和目标关系，如由GoogLe公司提出的DeepLab深度神经网络结构目前广泛应用于语义分割领域(L.-C.Chen,G.Papandreou,I.Kokkinos,K.Murphy,and A.L.Yuille.Deeplab:Semantic image segmentation with deep convolutional nets,atrous convolution,and fully connected crfs.arXiv:1606.00915,2016).但由于该通用语义分割网络计算实时性差，在嵌入式系统中难以应用。同时，语义分割也会来带边缘轮廓分割不明显、误检和漏检等情况。

[0004] 我们将语义分割应用在视觉SLAM系统语义建图中，从而使得所建立的空间网络地图中的每一个网络坐标点都具有高级的语义信息，让机器人对当前场景目标具有语义级理解，并且通过空间流形聚类算法优化语义分割所带来的误差，使得构建的语义地图更加准确。

发明内容

[0005] 本发明的目的是克服了上述现有技术的缺点，提供了一种将深度学习与视觉SLAM相结合、使机器人对场景目标具有语义级理解、减少语义分割的误差的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法及系统。

[0006] 为了实现上述目的，本发明的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法及系统如下：

[0007] 该基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法，其主要特点是，所述的方法包括以下步骤：

[0008] (1)嵌入式开发处理器通过RGB-D摄像头获取当前环境的彩色信息与深度信息；

[0009] (2)通过采集的图像得到特征点匹配对，并进行位姿估计，且获得场景空间点云数

据；

[0010] (3)利用深度学习对图像进行像素级语义分割，通过图像坐标系和世界坐标系映射，并使得空间点具有语义标注信息；

[0011] (4)通过流形聚类消除优化语义分割所带来的误差；

[0012] (5)进行语义建图，对空间点云进行拼接，得到由密集离散的点组成的点云语义地图。

[0013] 其中，所述的步骤(1)中的嵌入式处理器包括NVIDIA JETSON TX2系统。

[0014] 较佳地，所述的步骤(2)包括以下步骤：

[0015] (2.1)通过视觉SLAM技术提取图像特征点，进行特征匹配得到特征点匹配对；

[0016] (2.2)通过3D点对求解相机当前位姿；

[0017] (2.3)通过图优化Bundle Adjustment的方法进行更精确的位姿估计；

[0018] (2.4)通过回环检测消除帧间的累计误差，并获得场景空间点云数据。

[0019] 较佳地，所述的步骤(3)中的对图像进行像素级语义分割具体包括以下步骤：

[0020] (3.1)通过基于改进空洞卷积的GoogLeNet的特征提取层；

[0021] (3.2)通过基于改进空洞卷积的GoogLeNet的多尺度提取层；

[0022] (3.3)根据提取结果对图像进行分类。

[0023] 较佳地，所述的步骤(3.1)还包括特征提取层的设计过程，具体包括以下步骤：

[0024] (3.1.1)将GoogLeNet网络结构中Inception(3b)之后的最大池化层步长修改为1；

[0025] (3.1.2)将GoogLeNet网络结构中Inception(4a)、Inception(4b)、Inception(4c)、Inception(4d)、Inception(4e)部分使用空洞卷积代替，并设置空洞卷积为5×5且dilation为2的Pool；

[0026] (3.1.3)将GoogLeNet网络结构中Inception(4e)之后的最大池化层步长修改为1。

[0027] 较佳地，所述的步骤(3.2)还包括多尺度提取层的设计过程，具体包括以下步骤：

[0028] (3.2.1)基于空间金字塔池化进行多尺度处理；

[0029] (3.2.2)通过1×1卷积和不同采样率的空洞卷积提取不同尺度的特征图像；

[0030] (3.2.3)融合图像池化特征到模块中，将所述的特征图像经过1×1的卷积融合得到特征，并放入Softmax层进行像素点语义分类。

[0031] 较佳地，所述的步骤(4)具体包括以下步骤：

[0032] (4.1)计算空间点的切平面法向量；

[0033] (4.2)搜索未被分配类别的点x_i，判断是否所有点均已聚类，如果是，则继续步骤(4.5)；否则，x_i类别为c＝c+1，并创建空队列q；

[0034] (4.3)计算空间点x_i的切平面法向量v_i和距其小于0.01范围内所有点x_j的法向量v_j的夹角α_{ij}，判断是否存在$\alpha_{ij}<\sigma$或者$\alpha_{ij}>175°$，如果是，则x_j和x_i归为一类，x_j类别为c，并将满足条件的x_j压入队列q中；否则，继续步骤(4.4)；

[0035] (4.4)判断队列q是否非空，如果是，则令$x_i=q_1$，继续步骤(4.3)；否则继续步骤(4.1)；

[0036] (4.5)提取簇内点数最多的前k类点，剩下的点按照就近原则归类。

[0037] 较佳地，所述的步骤(4.1)中的计算空间点的切平面法向量，具体为：

[0038] 根据以下公式计算空间点的切平面法向量：

[0039] $$w = argmin_w \frac{1}{2} a w^T w = argmin_w \frac{1}{2} a,$$

[0040] 其中，$w \in R^{3\times1}$为该平面的单位法向量，a为特征值。

[0041] 较佳地，所述的步骤(5)包括以下步骤：

[0042] (5.1)根据RGB-D相机的精度特性，去除深度值太大或无效的点云；

[0043] (5.2)通过统计滤波器方法去除孤立的空间点，计算每个空间点与它最近N个空间点的距离均值，去除距离均值过大的空间点；

[0044] (5.3)通过空间网格原理，将所有空间点云填充进空间网格，使得每个空间网格只保留一个空间点。

[0045] 该基于上述方法的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的系统，其主要特点是，所述的系统包括：

[0046] 嵌入式开发处理器，用于构建视觉SLAM语义地图；

[0047] RGB-D相机，与所述的嵌入式开发处理器相连接，用于采集彩色数据和深度数据；

[0048] 建图程序，所述的建图程序在运行时根据深度学习与视觉SLAM，通过嵌入式开发处理器和RGB-D相机实现视觉SLAM语义建图，具体进行以下步骤处理：

[0049] (1)嵌入式开发处理器通过RGB-D摄像头获取当前环境的彩色信息与深度信息；

[0050] (2)通过采集的图像得到特征点匹配对，并进行位姿估计，且获得场景空间点云数据；

[0051] (3)利用深度学习对图像进行像素级语义分割，通过图像坐标系和世界坐标系映射，并使得空间点具有语义标注信息；

[0052] (4)通过流形聚类消除优化语义分割所带来的误差；

[0053] (5)进行语义建图，对空间点云进行拼接，得到由密集离散的点组成的点云语义地图。

[0054] 较佳地，所述的步骤(1)中的嵌入式处理器包括NVIDIA JETSON TX2系统。

[0055] 较佳地，所述的步骤(2)包括以下步骤：

[0056] (2.1)通过视觉SLAM技术提取图像特征点，进行特征匹配得到特征点匹配对；

[0057] (2.2)通过3D点对求解相机当前位姿；

[0058] (2.3)通过图优化Bundle Adjustment的方法进行更精确的位姿估计；

[0059] (2.4)通过回环检测消除帧间的累计误差，并获得场景空间点云数据。

[0060] 较佳地，所述的步骤(3)中的对图像进行像素级语义分割具体包括以下步骤：

[0061] (3.1)通过基于改进空洞卷积的GoogLeNet的特征提取层；

[0062] (3.2)通过基于改进空洞卷积的GoogLeNet的多尺度提取层；

[0063] (3.3)根据提取结果对图像进行分类。

[0064] 较佳地，所述的步骤(3.1)还包括特征提取层的设计过程，具体包括以下步骤：

[0065] (3.1.1)将GoogLeNet网络结构中Inception(3b)之后的最大池化层步长修改为1；

[0066] (3.1.2)将GoogLeNet网络结构中Inception(4a)、Inception(4b)、Inception(4c)、Inception(4d)、Inception(4e)部分使用空洞卷积代替，并设置空洞卷积为5×5且dilation为2的Pool；

[0067] (3.1.3)将GoogLeNet网络结构中Inception(4e)之后的最大池化层步长修改为1。

[0068] 较佳地，所述的步骤(3.2)还包括多尺度提取层的设计过程，具体包括以下步骤：

[0069] (3.2.1)基于空间金字塔池化进行多尺度处理；

[0070] (3.2.2)通过1×1卷积和不同采样率的空洞卷积提取不同尺度的特征图像；

[0071] (3.2.3)融合图像池化特征到模块中，将所述的特征图像经过1×1的卷积融合得到特征，并放入Softmax层进行像素点语义分类。

[0072] 较佳地，所述的步骤(4)具体包括以下步骤：

[0073] (4.1)计算空间点的切平面法向量；

[0074] (4.2)搜索未被分配类别的点x_i，判断是否所有点均已聚类，如果是，则继续步骤(4.5)；否则，x_i类别为c＝c+1，并创建空队列q；

[0075] (4.3)计算空间点x_i的切平面法向量v_i和距其小于0.01范围内所有点x_j的法向量v_j的夹角α_{ij}，判断是否存在$\alpha_{ij}<\sigma$或者$\alpha_{ij}>175°$，如果是，则x_j和x_i归为一类，x_j类别为c，并将满足条件的x_j压入队列q中；否则，继续步骤(4.4)；

[0076] (4.4)判断队列q是否非空，如果是，则令$x_i=q_1$，继续步骤(4.3)；否则继续步骤(4.1)；

[0077] (4.5)提取簇内点数最多的前k类点，剩下的点按照就近原则归类。

[0078] 较佳地，所述的步骤(4.1)中的计算空间点的切平面法向量，具体为：

[0079] 根据以下公式计算空间点的切平面法向量：

[0080] $$w = argmin_w \frac{1}{2} a w^T w = argmin_w \frac{1}{2} a,$$

[0081] 其中，$w \in R^{3\times1}$为该平面的单位法向量，a为特征值。

[0082] 较佳地，所述的步骤(5)包括以下步骤：

[0083] (5.1)根据RGB-D相机的精度特性，去除深度值太大或无效的点云；

[0084] (5.2)通过统计滤波器方法去除孤立的空间点，计算每个空间点与它最近N个空间点的距离均值，去除距离均值过大的空间点；

[0085] (5.3)通过空间网格原理，将所有空间点云填充进空间网格，使得每个空间网格只保留一个空间点。

[0086] 采用了本发明的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法及系统，系统采用嵌入式开发处理器，通过对RGB-D相机采集到的彩色数据和深度数据，利用视觉SLAM技术，提取图像特征点，进行特征匹配，再利用Bundle Adjustment的方法得到更精确的机器人位姿估计，使用回环检测消除帧间的累计误差。在获得机器人实时定位信息的同时，采用了一种针对GoogLeNet深度神经网络的空洞卷积设计方法，利用改进深度神经网络实现实时语义分割的特征提取，将语义分割结果结合视觉SLAM系统得到语义级的建图。并通过流形聚类消除优化语义分割所带来的误差，通过八叉树建图后，空间网络地图具有更高级的语义信息，并且构建出的语义地图更加准确。网络的改进提升了系统的实时处理能力，本方法和系统的语义分割网络在NVIDIA Jetson TX2台上的时间消耗为0.099s/幅，符合在实时建图过程中的使用需求。

附图说明

[0087] 图1为本发明的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法的流程图。

[0088] 图2为本发明的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法的语义分割流程图。

[0089] 图3为本发明的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法的空洞卷积示意图。

[0090] 图4为本发明的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法的实验结果示意图。

[0091] 图5为本发明的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法及系统的NVIDIA Jetson TX2处理器示意图。

具体实施方式

[0092] 为了能够更清楚地描述本发明的技术内容,下面结合具体实施例来进行进一步的描述。

[0093] 该基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法,其中,所述的方法包括以下步骤:

[0094] (1)嵌入式开发处理器通过RGB-D摄像头获取当前环境的彩色信息与深度信息;

[0095] (2)通过采集的图像得到特征点匹配对,并进行位姿估计,且获得场景空间点云数据;

[0096] (2.1)通过视觉SLAM技术提取图像特征点,进行特征匹配得到特征点匹配对;

[0097] (2.2)通过3D点对求解相机当前位姿;

[0098] (2.3)通过图优化Bundle Adjustment的方法进行更精确的位姿估计;

[0099] (2.4)通过回环检测消除帧间的累计误差,并获得场景空间点云数据;

[0100] (3)利用深度学习对图像进行像素级语义分割,通过图像坐标系和世界坐标系映射,并使得空间点具有语义标注信息;

[0101] (3.1)通过基于改进空洞卷积的GoogLeNet的特征提取层;

[0102] (3.1.1)将GoogLeNet网络结构中Inception(3b)之后的最大池化层步长修改为1;

[0103] (3.1.2)将GoogLeNet网络结构中Inception(4a)、Inception(4b)、Inception(4c)、Inception(4d)、Inception(4e)部分使用空洞卷积代替,并设置空洞卷积为5×5且dilation为2的Pool;

[0104] (3.1.3)将GoogLeNet网络结构中Inception(4e)之后的最大池化层步长修改为1;

[0105] (3.2)通过基于改进空洞卷积的GoogLeNet的多尺度提取层;

[0106] (3.2.1)基于空间金字塔池化进行多尺度处理;

[0107] (3.2.2)通过1×1卷积和不同采样率的空洞卷积提取不同尺度的特征图像;

[0108] (3.2.3)融合图像池化特征到模块中,将所述的特征图像经过1×1的卷积融合得到特征,并放入Softmax层进行像素点语义分类;

[0109] (3.3)根据提取结果对图像进行分类;

[0110] (4)通过流形聚类消除优化语义分割所带来的误差;

[0111] (4.1)计算空间点的切平面法向量;

[0112] (4.2)搜索未被分配类别的点x_i,判断是否所有点均已聚类,如果是,则继续步骤(4.5);否则,x_i类别为c=c+1,并创建空队列q;

[0113] (4.3)计算空间点x_i的切平面法向量v_i和距其小于0.01范围内所有点x_j的法向量v_j的夹角α_{ij},判断是否存在$\alpha_{ij}<\sigma$或者$\alpha_{ij}>175°$,如果是,则x_j和x_i归为一类,x_j类别为c,并将满足条件的x_j压入队列q中;否则,继续步骤(4.4);

[0114] (4.4)判断队列q是否非空,如果是,则令$x_i=q_1$,继续步骤(4.3);否则继续步骤(4.1);

[0115] (4.5)提取簇内点数最多的前k类点,剩下的点按照就近原则归类;

[0116] (5)进行语义建图,对空间点云进行拼接,得到由密集离散的点组成的点云语义地图;

[0117] (5.1)根据RGB-D相机的精度特性,去除深度值太大或无效的点云;

[0118] (5.2)通过统计滤波器方法去除孤立的空间点,计算每个空间点与它最近N个空间点的距离均值,去除距离均值过大的空间点;

[0119] (5.3)通过空间网格原理,将所有空间点云填充进空间网格,使得每个空间网格只保留一个空间点。

[0120] 作为本发明的优选实施方式,所述的步骤(1)中的嵌入式处理器包括NVIDIAJETSON TX2系统。

[0121] 作为本发明的优选实施方式,所述的步骤(4.1)中的计算空间点的切平面法向量,具体为:

[0122] 根据以下公式计算空间点的切平面法向量:

[0123] $$w = argmin_w \frac{1}{2} a w^T w = argmin_w \frac{1}{2} a,$$

[0124] 其中,$w\in R^{3\times1}$为该平面的单位法向量,a为特征值。

[0125] 该基于上述方法的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的系统,其中,所述的系统包括:

[0126] 嵌入式开发处理器,用于构建视觉SLAM语义地图;

[0127] RGB-D相机,与所述的嵌入式开发处理器相连接,用于采集彩色数据和深度数据;

[0128] 建图程序,所述的建图程序在运行时根据深度学习与视觉SLAM,通过嵌入式开发处理器和RGB-D相机实现视觉SLAM语义建图,具体进行以下步骤处理:

[0129] (1)嵌入式开发处理器通过RGB-D摄像头获取当前环境的彩色信息与深度信息;

[0130] (2)通过采集的图像得到特征点匹配对,并进行位姿估计,且获得场景空间点云数据;

[0131] (2.1)通过视觉SLAM技术提取图像特征点,进行特征匹配得到特征点匹配对;

[0132] (2.2)通过3D点对求解相机当前位姿;

[0133] (2.3)通过图优化Bundle Adjustment的方法进行更精确的位姿估计;

[0134] (2.4)通过回环检测消除帧间的累计误差,并获得场景空间点云数据;

[0135] (3)利用深度学习对图像进行像素级语义分割,通过图像坐标系和世界坐标系映射,并使得空间点具有语义标注信息;

[0136] (3.1)通过基于改进空洞卷积的GoogLeNet的特征提取层;

[0137] (3.1.1)将GoogLeNet网络结构中Inception(3b)之后的最大池化层步长修改为1;

[0138] (3.1.2)将GoogLeNet网络结构中Inception(4a)、Inception(4b)、Inception

(4c)、Inception(4d)、Inception(4e)部分使用空洞卷积代替，并设置空洞卷积为5×5且dilation为2的Pool；

[0139] (3.1.3)将GoogLeNet网络结构中Inception(4e)之后的最大池化层步长修改为1；

[0140] (3.2)通过基于改进空洞卷积的GoogLeNet的多尺度提取层；

[0141] (3.2.1)基于空间金字塔池化进行多尺度处理；

[0142] (3.2.2)通过1×1卷积和不同采样率的空洞卷积提取不同尺度的特征图像；

[0143] (3.2.3)融合图像池化特征到模块中，将所述的特征图像经过1×1的卷积融合得到特征，并放入Softmax层进行像素点语义分类；

[0144] (3.3)根据提取结果对图像进行分类；

[0145] (4)通过流形聚类消除优化语义分割所带来的误差；

[0146] (4.1)计算空间点的切平面法向量；

[0147] (4.2)搜索未被分配类别的点x_i，判断是否所有点均已聚类，如果是，则继续步骤(4.5)；否则，x_i类别为c＝c+1，并创建空队列q；

[0148] (4.3)计算空间点x_i的切平面法向量v_i和距其小于0.01范围内所有点x_j的法向量v_j的夹角α_{ij}，判断是否存在$\alpha_{ij}<\sigma$或者$\alpha_{ij}>175°$，如果是，则x_j和x_i归为一类，x_j类别为c，并将满足条件的x_j压入队列q中；否则，继续步骤(4.4)；

[0149] (4.4)判断队列q是否非空，如果是，则令$x_i=q_1$，继续步骤(4.3)；否则继续步骤(4.1)；

[0150] (4.5)提取簇内点数最多的前k类点，剩下的点按照就近原则归类；

[0151] (5)进行语义建图，对空间点云进行拼接，得到由密集离散的点组成的点云语义地图；

[0152] (5.1)根据RGB-D相机的精度特性，去除深度值太大或无效的点云；

[0153] (5.2)通过统计滤波器方法去除孤立的空间点，计算每个空间点与它最近N个空间点的距离均值，去除距离均值过大的空间点；

[0154] (5.3)通过空间网格原理，将所有空间点云填充进空间网格，使得每个空间网格只保留一个空间点。

[0155] 作为本发明的优选实施方式，所述的步骤(1)中的嵌入式处理器包括NVIDIA JETSON TX2系统。

[0156] 作为本发明的优选实施方式，所述的步骤(4.1)中的计算空间点的切平面法向量，具体为：

[0157] 根据以下公式计算空间点的切平面法向量：

[0158] $$w = argmin_w \frac{1}{2} aw^T w = argmin_w \frac{1}{2} a,$$

[0159] 其中，$w \in R^{3\times1}$为该平面的单位法向量，a为特征值。

[0160] 本发明的具体实施方式中，本发明涉及无人机器人系统实时定位与建图的技术领域，是一种基于空洞卷积深度神经网络的视觉SLAM语义建图方法及系统。系统采用嵌入式开发处理器，通过对RGB-D相机采集到的彩色数据和深度数据，利用视觉SLAM技术，提取图像特征点，进行特征匹配，再利用Bundle Adjustment的方法得到更精确的机器人位姿估计，使用回环检测消除帧间的累计误差。在获得机器人实时定位信息的同时，采用了一种针对GoogLeNet深度神经网络的空洞卷积设计方法，利用改进深度神经网络实现实时语义分

割的特征提取，将语义分割结果结合视觉SLAM系统得到语义级的建图，并通过流形聚类消除优化语义分割所带来的误差，通过八叉树建图后，空间网络地图具有更高级的语义信息，并且构建出的语义地图更加准确。

[0161] 该基于上述系统基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法，其中，包括以下步骤：

[0162] (1)采用嵌入式开发处理器，通过RGB-D摄像头获取当前环境的彩色信息与深度信息；

[0163] (2)对由相机采集得到的图像，利用视觉SLAM技术，提取图像特征点，进行特征匹配得到特征点匹配对；利用3D点对求解相机当前位姿；利用图优化Bundle Adjustment的方法进行更精确的位姿估计；使用回环检测消除帧间的累计误差，并获得场景空间点云数据；

[0164] (3)利用深度学习对图像进行像素级语义分割，利用图像坐标系和世界坐标系的关系映射到空间中，使得每一个空间点都具有语义标注信息；

[0165] (4)采用流形聚类优化语义分割带来的误差；

[0166] (5)进行语义建图，对空间点云进行拼接，最终得到一个由密集离散的点组成的点云语义地图。

[0167] 在上述实例中，所述的步骤(1)所述的嵌入式处理器包括NVIDIA Jetson TX2系统及同类设备。

[0168] 在上述实例中，所述的步骤(2)中采用通用视觉SLAM及其局部改进技术。

[0169] 在上述实例中，所述的步骤(3)中语义分割网络具体包括以下结构：

[0170] (31)特征提取层；

[0171] (32)多尺度提取层；

[0172] (33)分类层；

[0173] 在上述实例中，所述的步骤(31)中所述的特征提取层具体包括以下结构：

[0174] (311)采用GoogLeNet网络结构作为DeepLab模型的前端特征提取层；

[0175] (312)将GoogLeNet网络结构中Inception(3b)之后的最大池化层步长修改为1，从而扩大了特征尺度，保证输出分辨率不变；

[0176] (313)将GoogLeNet网络结构中Inception(4a)部分使用空洞卷积代替，设置dilation为2，5×5的Pool，从而扩大特征尺度；

[0177] (314)将GoogLeNet网络结构中Inception(4b)部分使用空洞卷积代替，设置dilation为2，5×5的Pool，从而扩大特征尺度；

[0178] (315)将GoogLeNet网络结构中Inception(4c)部分使用空洞卷积代替，设置dilation为2，5×5的Pool，从而扩大特征尺度；

[0179] (316)将GoogLeNet网络结构中Inception(4d)部分使用空洞卷积代替，设置dilation为2，5×5的Pool，从而扩大特征尺度；

[0180] (317)将GoogLeNet网络结构中Inception(4e)部分使用空洞卷积代替，设置dilation为2，5×5的Pool，从而扩大特征尺度；

[0181] (316)将GoogLeNet网络结构中Inception(4e)之后的最大池化层步长修改为1，从而扩大了特征尺度，保证输出分辨率不变；

[0182] 其中原始的GoogLeNet输入尺寸为224，特征输出尺寸为7，相当于缩小了32倍，将

后两层池化层的步长修改为1，并将原有的普通卷积修改为空洞卷积，这样对于输入尺寸为321，特征图的输出尺寸为41，相当于缩小了8倍，从而扩大了特征尺度。

[0183] 在上述实例中，所述的步骤(32)中所述的多尺度层具体包括以下结构：

[0184] (321)基于空间金字塔池化地方式进行多尺度处理；

[0185] (322)对空间金字塔池化模型进行优化，使用1×1卷积以及不同采样率(6、12、18)的空洞卷积提取不同尺度感受野的特征；

[0186] (323)将图像池化特征融合到模块中，然后将所得到的特征图像都经过1×1的卷积后融合(Concat)得到最后的特征，再放入Softmax层进行像素点语义分类。。

[0187] 在上述实例中，所述地步骤(4)中所述的流形聚类具体包括以下步骤：

[0188] (41)计算每一个空间点的切平面法向量，设当前聚类类别c＝0；

[0189] (42)搜索一个还没有被分配类别的点x_i，如果所有点均已聚类，则执行步骤(85)，否则，设x_i类别为c＝c+1，并且创建一个空的队列q；

[0190] (43)计算空间点x_i的切平面法向量v_i和其距离小于0.01范围内所有点x_j的法向量v_j的夹角α_{ij}，如果$\alpha_{ij}<\sigma$或者$\alpha_{ij}>175°$，那么x_j和x_i归为一类，x_j类别为c，并将满足条件的x_j压入队列q中；

[0191] (44)如果队列q非空，则令$x_i=q_1$，继续执行第3步，否则跳转到第1步；

[0192] (45)提取簇内点数最多的前k类点，剩下的点按照就近原则归类。

[0193] 其中，步骤(41)中切平面法向量的计算步骤为：

[0194] 设n个空间点组成矩阵$X \in \mathbb{R}^{3\times n}$，X的协方差矩阵$\Sigma=E[(X-\mu)(X-\mu)^T]$

[0195] 设$w\in R^{3\times 1}$为这个平面的单位法向量，$Z=w^TX$为这n个点在这个单位法向量上的投影长度，建立模型：

[0196] $$w = \underset{w}{argmin}\frac{1}{2}w^T\Sigma w$$

[0197] $s.t.w^Tw=1$

[0198] 利用拉格朗日乘子法求解：

[0199] $$\mathcal{L}(w,a) = \frac{1}{2}[w^T\Sigma w - a(w^Tw-1)]$$

[0200] 对上式求偏导数得：

[0201] $$\begin{cases}\frac{\partial \mathcal{L}}{\partial w} = \Sigma w - aw = 0\\ \frac{\partial \mathcal{L}}{\partial a} = w^Tw - 1 = 0\end{cases}$$

[0202] w需要单位化，上式中a对应特征值，即有$w = argmin_w \frac{1}{2}aw^Tw = argmin_w \frac{1}{2}a$并且协方差矩阵是半正定矩阵，所以空间向量w为协方差矩阵的Σ多对应特征值最小的单位特征向量。

[0203] 在上述实例中，所述的步骤(5)中所述的建图算法具体包括以下步骤：

[0204] (51)生成每一帧点云信息时，根据RGB-D相机的精度特性，去除深度值太大或者无效的点云；

[0205] (52)采用统计滤波器方法去除孤立的空间点，计算每个空间点与它最近N个空间

点的距离均值，去除距离均值过大的空间点，从而保留密集空间点，去掉了孤立的噪声点；

[0206] (53)利用空间网格原理，将所有的空间点云填充进空间网格中，保证每个空间网格仅只保留一个空间点，相当于对空间点云进行降采样，从而节省了很多存储空间。

[0207] 其中，使用八叉树数据结构建立空间网络地图。

[0208] 对于一个空间立方体，将其分为八个区域，相同的，每个子区域继续分割成八个区域，这样动态的创建一棵八叉树地图。

[0209] 下面结合附图及具体实施例详细介绍，本发明的基于空洞卷积神经网络的视觉SLAM语义建图方法。

[0210] 基于空洞卷积神经网络的视觉SLAM语义建图方法及系统流程如图1所示：

[0211] 由RGB-D摄像头采集的图像数据，挑选相似度不高的帧作为关键帧，关键帧包含彩色图像，深度图像和当前位姿，对彩色图像进行语义分割，首先通过使用基于改进空洞卷积的GoogLeNet的特征提取层，多尺度层，得到原始语义点云。对原始语义点云进行滤波操作，结合深度图像进行流行聚类，最终结合位姿信息一起进行八叉树建图，网络的改进提升了系统的实时处理能力，能够在基于NVIDIAJETSON TX2的嵌入式平台上实时实现。

[0212] 基于空洞卷积神经网络的视觉SLAM语义建图方法及系统流程中，通过深度学习语义分割网络来获取图像的语义信息，系统流程如图2所示，主要分为特征提取，多尺度提取和分类三个部分。

[0213] 基于空洞卷积神经网络的视觉SLAM语义建图方法及系统流程中使用的空洞卷积如图3所示：

[0214] 将卷积和池化视作同种操作，假设中间紫色点部分作为输入，图的绿色部分为普通卷积过程，经过步长分别为2、1、2、1的卷积(或池化)过程后，得到特征。最上层的特征点所对应的感受野为整个输入层。

[0215] 为了扩大特征尺寸，使用空洞卷积，为图中粉色部分，将步长全部改为1，第一层卷积步长改变后，令dilation为1，得到的特征数目扩大了两倍，在进行第二层卷积操作时，令dilation为2，即做卷积操作时，间隔1个点与卷积核卷积，得到特征还是原来普通卷积的两倍，且特征点的感受野不变，继续进行第三层卷积操作，同将步长改为1，为了保持相同的感受野，此时的dilation同样应该为2。在第四层卷积操作时，此时dilation要为4才能保持感受野不变。

[0216] 在使用空洞卷积时需要注意：

[0217] s1.在上一层卷积操作的步长由$stride_{old}$变为$stride_{new}$，为了保持感受野不变，接下来所有的卷积层操作都要进行空洞率为$\frac{stride_{old}}{stride_{new}}$的带孔卷积；

[0218] s2.当前层空洞卷积操作的空洞率如以下公式。

[0219] $$\text{dilation} = \frac{stride_{old}^{1}}{stride_{new}^{1}} \times \frac{stride_{old}^{2}}{stride_{new}^{2}} \cdots\cdots \frac{stride_{old}^{N}}{stride_{new}^{N}}$$

[0220] 其中N代表之前层步长改变次数，$stride_{old}^{N}/stride_{new}^{N}$为第N次步长的改变。

[0221] 基于空洞卷积神经网络的视觉SLAM语义建图结果如图4所示。图中图像是在两个场景中实验的结果，左为办公室场景，右为实验室场景。图中第一行为本系统输出的具有语义信息的建图结果，其中椅子、人、植物分别用红色、粉色、绿色标示；第二行为传统视觉

SLAM建立的无语义信息的建图结果。实验结果表明，本发明能使机器人很好的理解当前场景中的主要目标。本发明所涉及的软件及算法均在NVIDIA Jetson TX2嵌入式平台上事项，其处理器图示如图5所示。

[0222] 采用了本发明的基于空洞卷积深度神经网络实现视觉SLAM语义建图功能的方法及系统，系统采用嵌入式开发处理器，通过对RGB-D相机采集到的彩色数据和深度数据，利用视觉SLAM技术，提取图像特征点，进行特征匹配，再利用Bundle Adjustment的方法得到更精确的机器人位姿估计，使用回环检测消除帧间的累计误差。在获得机器人实时定位信息的同时，采用了一种针对GoogLeNet深度神经网络的空洞卷积设计方法，利用改进深度神经网络实现实时语义分割的特征提取，将语义分割结果结合视觉SLAM系统得到语义级的建图。并通过流形聚类消除优化语义分割所带来的误差，通过八叉树建图后，空间网络地图具有更高级的语义信息，并且构建出的语义地图更加准确。网络的改进提升了系统的实时处理能力，本方法和系统的语义分割网络在NVIDIA Jetson TX2台上的时间消耗为0.099s/幅，符合在实时建图过程中的使用需求。

[0223] 在此说明书中，本发明已参照其特定的实施例作了描述。但是，很显然仍可以作出各种修改和变换而不背离本发明的精神和范围。因此，说明书和附图应被认为是说明性的而非限制性的。

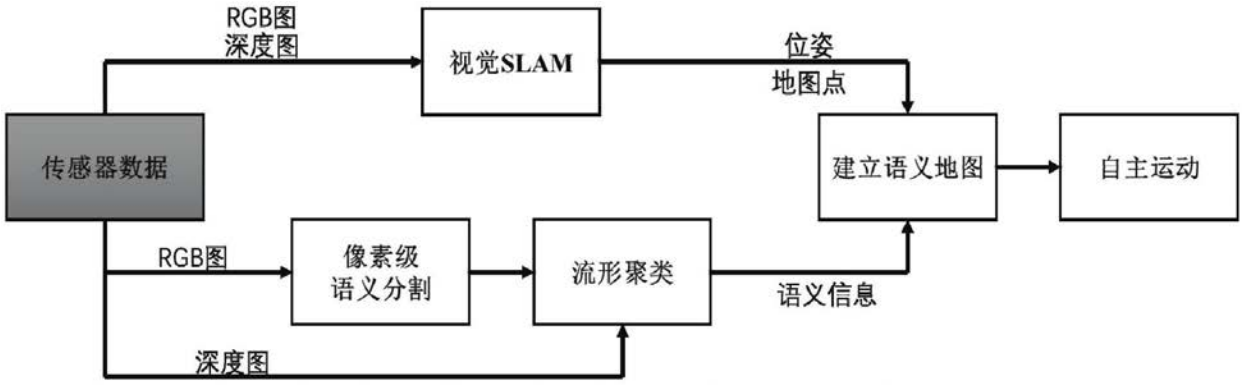

RGB图
深度图
视觉SLAM
位姿
地图点
传感器数据
建立语义地图
自主运动
RGB图
像素级
语义分割
流形聚类
语义信息
深度图

图1


RGB图像
输入
特征提取层
多尺度层
分类层
语义信息

图2

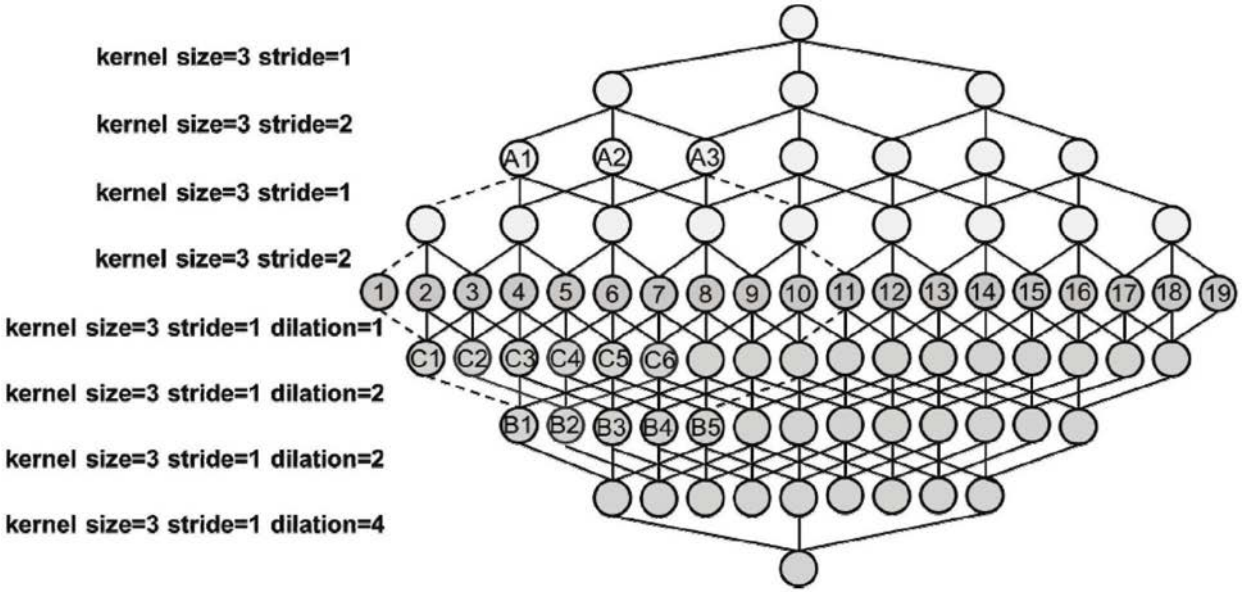

kernel size=3 stride=1
kernel size=3 stride=2
kernel size=3 stride=1
kernel size=3 stride=2
kernel size=3 stride=1 dilation=1
kernel size=3 stride=1 dilation=2
kernel size=3 stride=1 dilation=2
kernel size=3 stride=1 dilation=4
A1
A2
A3
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
C1
C2
C3
C4
C5
C6
B1
B2
B3
B4
B5

图3

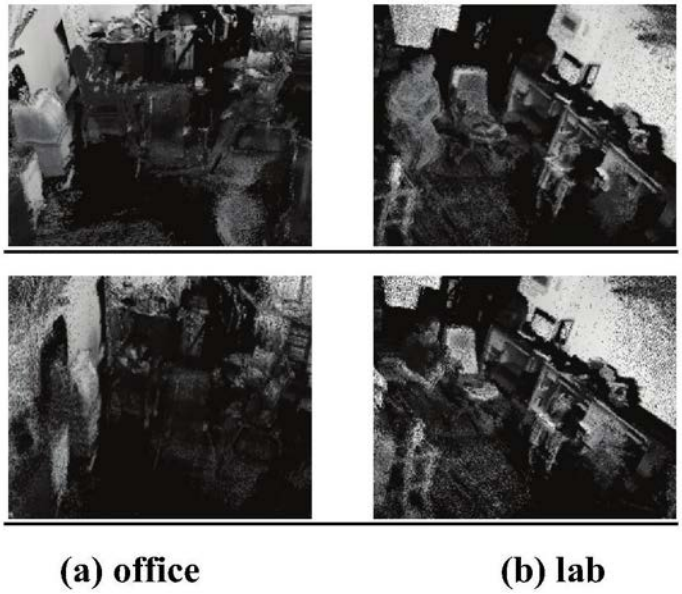


图4

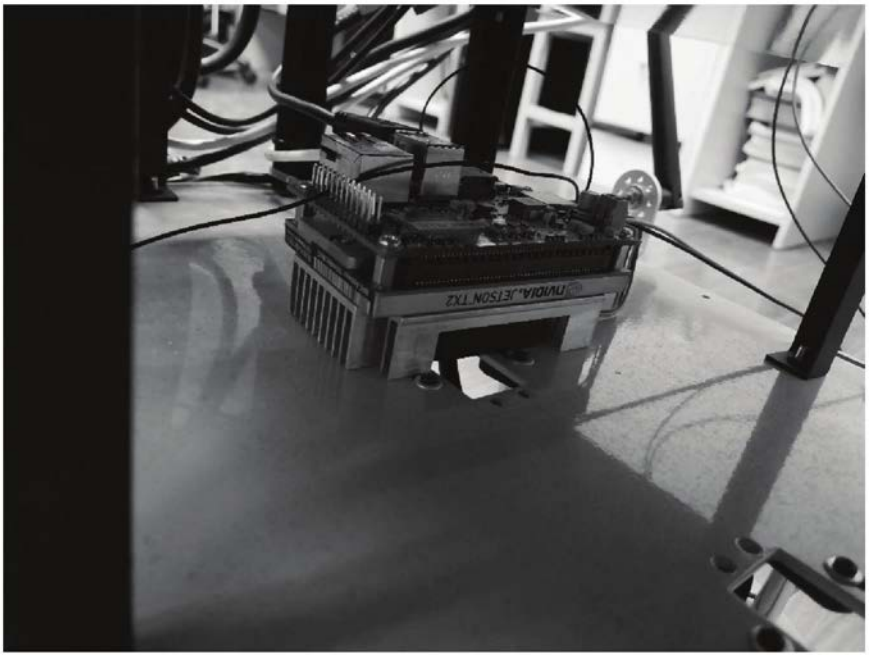


图5