第13周汇报

——张溢炉

1. 《Mapping Technology in Visual SLAM: A Review》总结
2. 内容摘要

本文主要介绍一些传统的建图方法

1. 总结及理解

* 对于移动机器人来说，SLAM作为底层技术，可以用于上层定位、导航、避障、重建和可视化等，稀疏地图可以用于机器人定位，但导航、避障和重建需要用到稠密地图。
* 视觉SLAM可以用于建立稀疏、半稠密和稠密地图。
* 基于特征的方法主要用于建立稀疏地图，Mur-Artal [9]提出了一种建立半稠密地图的方法。直接法可以用于建立半稠密和稠密地图，对环境特征和图像质量不太敏感。稠密地图可以用建立网格地图和八叉树地图，用于导航和避障
* ORB-SLAM的各个优点，它可以用于室内和室外环境：

1）新三维点的匹配更高效

2)闭环检测可以解决长期漂移积累误差

3)目标函数相位可以非常有效地求解

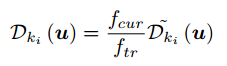
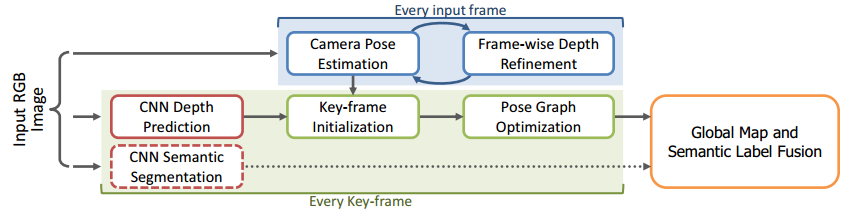
* 建立稀疏地图计算量较小，可以使用板上的处理器。但建立半稠密和稠密地图计算量较大，通常需要用GPU
* 未来的发展方向和应用有3D实时建图、建立网格地图和拓扑地图、增强现实AR、医疗应用

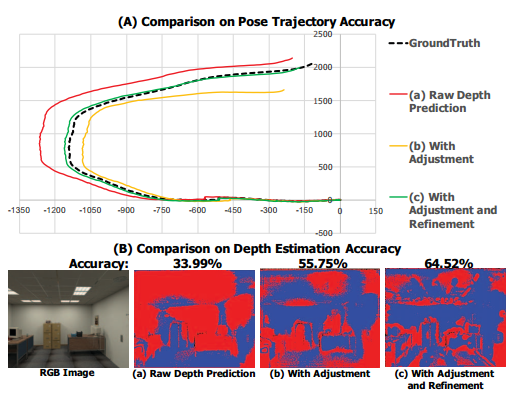
1. 《CNN-SLAM: Real-time dense monocular SLAM with learned depth prediction》总结
2. 内容摘要

本文介绍了将CNN与半稠密SLAM方法结合的单目相机稠密建图方法，另外还在图中加入了语义信息

1. 总结及理解

* 动机：克服单目相机的一些局限，如绝对尺寸不确定性、旋转和低纹理区域遇到的问题
* 模型如下图所示，模型中有三个通道，中间的是CNN通道，可以通过预训练好的CNN模型生成全局的初始化深度，CNN模型建立的深度在物体边缘比较模糊；这正好与半稠密方法互补，故再利用最上层的半稠密方法来精细化物体的边缘；最下层通道与上两层相互独立，增加语义信息有助于智能体对周围环境的理解和相互作用。注意，因为实验使用的摄像头与采集数据集使用的摄像头不同，其内参也会不一样，所以使用CNN预测的深度需要经过如下公式调整



* 下图展现的是调整CNN预测参数（with adjustment）和精细化（Refinement）带来的效果提升
* 文中使用的两个数据集是：the ICL-NUIM dataset [8] and TUM RGB-D SLAMdataset [26]
* 语义标记中常用的四个标记是：floor（如地板）,vertical structure（如墙）, large structure（如家具）, small structure（如杯子）
* 实验时，CNN和语义分割模型是在GPU运行的，其他运算在CPU上运行。目前来看深度学习的计算量太大，需要使用到CPU和GPU，不太适合计算能力弱、实时性要求高的小型移动机器人，但5G等通讯技术在速度上的提升，可以考虑将数据传输到云端处理。此外，像汽车这种大型的移动机器人，可以搭载GPU使用深度学习来做SLAM，有利于提高自动驾驶的安全性
* 未解决问题：具体公式还需在理解一些基础公式（如g2o、Huber function等）后来理解

1. 《Past, Present, and Future of Simultaneous Localization and Mapping: Toward the

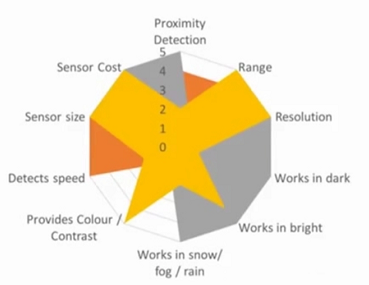
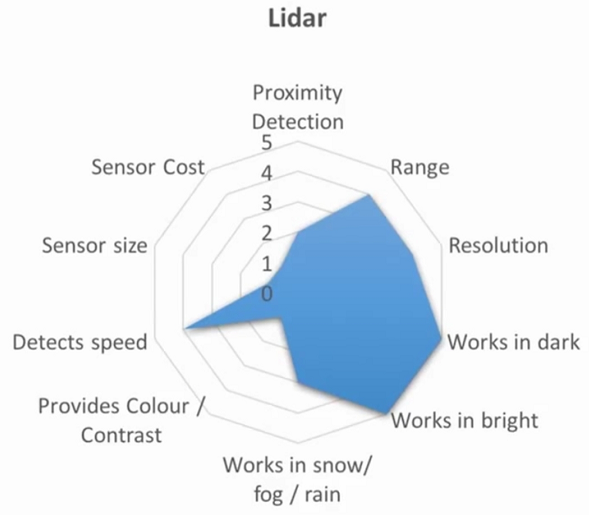
Robust-Perception Age》总结

1. 内容摘要

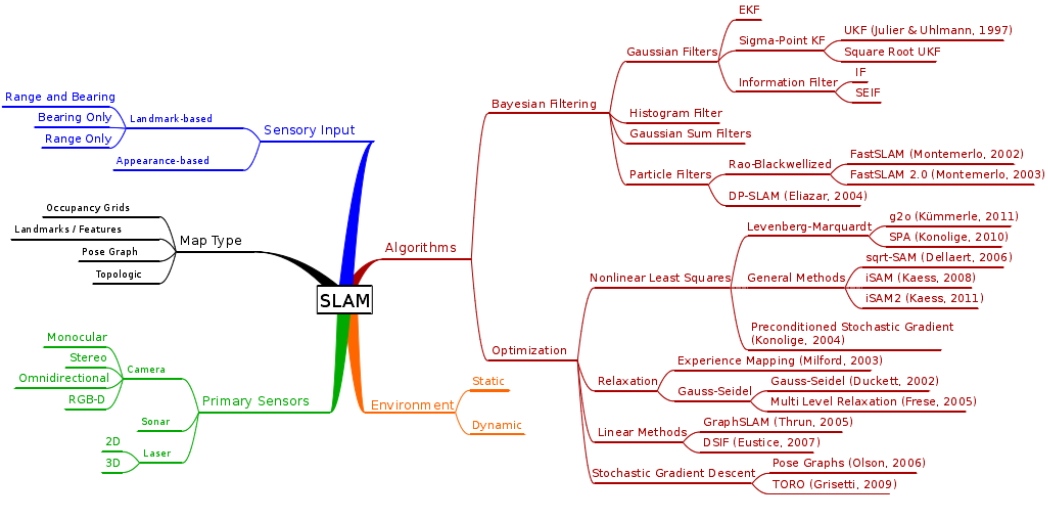
本文较上周两篇对SLAM过去30年的发展做了比较完整的回顾，不仅仅是视觉SLAM。可将SLAM的发展分为2个时代：古典时代（1986-2004）、算法分析时代（2004-2015）

1. 总结及理解

* 建图主要有两个作用：第一，给上层算法使用；第二，可以减小视觉里程计估计的误差
* 以下分别是超声波、摄像头、雷达、激光的优缺点图，单个传感器有一些局限，融合多种传感器可以取长补短，适应不同环境



* visual-inertial navigation (VIN)方法融合了相机和IMU，不用做闭环检测，取长补短。相机适合长时间移动测位姿，但相机不能判断是物体动还是相机动。IMU擅长快速移动情况，可以用于判断相机的短时间的移动
* 状态估计的两种方法
* 历史上很长一段时间，研究者们使用滤波器，尤其是扩展卡尔曼滤波器（EKF），只关心当前状态
* 近年来普遍使用非线性优化方法，使用所有时刻采集的数据进行状态估计，被认为优于滤波器，而成为主流方法
* SLAM的常见公式有：maximum a posteriori estimation, factor graph optimization, graph-SLAM, full smoothing, smoothing and mapping(SAM)
* SLAM的分类如下图，从算法、环境、传感器、地图类型、传感器的输入进行了分类

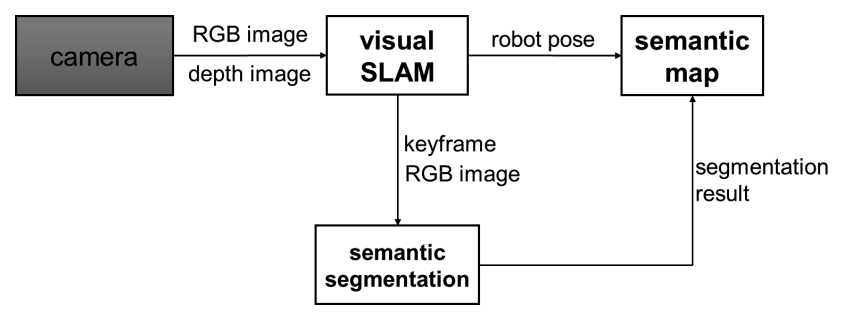


1. 《Real-Time Semantic Mapping of Visual SLAM Based on DCNN》总结
2. 内容摘要

本文提出了一种使用DCNN模型建立语义地图，并具有较好的实时性

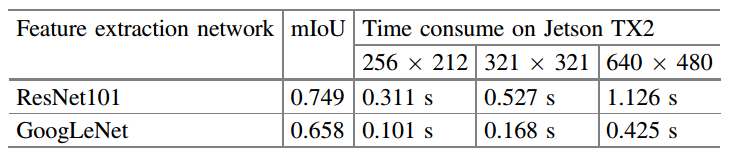
1. 总结及理解

* 动机：传统的SLAM只包含距离和颜色等低级信息，本文使用深度学习在地图中加入语义高级信息，并保持SLAM的实时性，促进机器人对陌生环境的理解，加强人机交互
* 本文的整体模型如下，首先使用ORB-SLAM选择关键帧，实时定位和建图，然后将关键帧通过语义分割模型进行处理，得到语义图，最后将ORB-SLAM建立的3D图与语义图合成。



* 语义分割模型如下，包括feature extraction、muti-scale、classification三个模块
* feature extraction：使用了一个1\*1的卷积层将64个通道减少到32个通道加速后面运算；使用dilated convolution卷积提高输出特征的分辨率
* muti-scale：先使用图像金字塔方法池化，再使用4个dilation为6, 12, 18，24及核为3\*3的dilated convolution并行处理，最后将结果串联得到最后特征
* classification：使用softmax进行分类



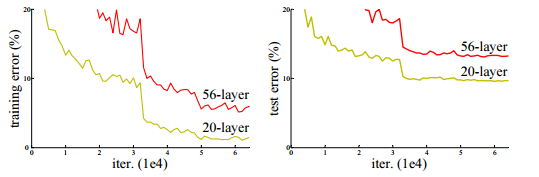
* 本文使用了NVIDIA Jetson TX2嵌入式模块来实现视觉SLAM的实时处理。它是一台模块化 AI 超级计算机，采用 NVIDIA Pascal™ 架构，具有 256 个CUDA核心。更棒的是，它性能强大，但外形小巧，节能高效，非常适合机器人、无人机、智能摄像机和便携医疗设备等智能边缘设备。它支持 Jetson TX1 模块的所有功能，同时可以铸就更大型、更复杂的深度神经网络。
* 实验结果如下图所示，虽然准确率相对ResNet101低0.091，但在各种分辨率下速度是ResNet101的3倍左右。从本文实验结果来看，将深度学习用于小型移动机器人具有一定可行性

1. 《Deep Residual Learning for Image Recognition》总结
2. 内容摘要

本文是ILSVRC 2015分类任务的冠军模型：ResNet

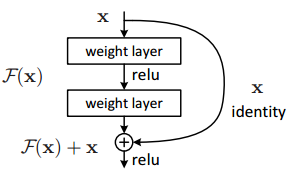
1. 总结及理解

* 动机：一般来说，随着模型深度增加，在不过拟合的情况下，训练误差和测试误差会更小，但是实验结果并不是这样，刚刚相反，出现了退化现象，实验结果如下图所示。ResNet旨在解决这个问题。假设：如果最优函数更接近于恒等映射而不是零映射，那么solvers应该更容易找到与恒等映射相关的扰动，而不是将函数作为一个新的函数来学习，这样学习更快。

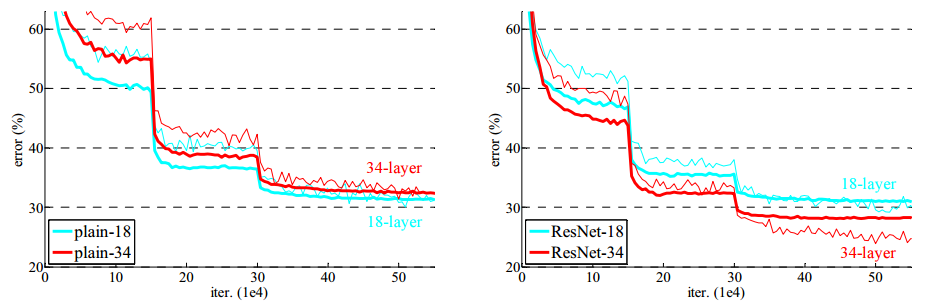


* ResNet由堆叠的一个个小模块（如下图所示）组成，相对VGG的改变，ResNet的非线性层拟合的目标函数是F(x) := H(x) – x，H(x) := F(x) – x是任务待拟合的函数。其中X的映射有三种方式，从实验结果来看A\B\C对实验结果影响不大，但计算量逐个增大：

1. 当非线性层输出维度和x相同时，对应元素直接相加；不同的话，填充0
2. 当非线性层输出维度和x相同时，对应元素直接相加；不同的话，x进行线性变换，然后与F(x)对应元素相加
3. 任何情况下，x都进行线性变换，然后与F(x)对应元素相加



* 下图为plain网络和ResNet实验结果，细线为训练误差，粗线为测试误差。ResNet-34的训练和测试误差都小于ResNet-18，说明ResNet解决了退化的问题



* 待解决问题：batch normalization (BN) [16]