DynaSLAM

- **论文：** DynaSLAM_Tracking_Mapping_and_Inpainting_in_Dynamic_Scenes.pdf
- **开源：** https://github.com/BertaBescos/DynaSLAM

在SLAM算法中的一个典型假设是场景固定。这样的一个强假设限制了大多数视觉SLAM系统在有人居住下真实环境中的使用，而这正是一些诸如服务机器人、自动驾驶车辆等相关应用的目标场景。本文中，作者提出DynaSLAM系统，这是一种建立在ORB-SLAM2上的视觉SLAM系统，同时增加了动态物体检测和背景修复功能。DynaSLAM在单目、立体、RGB-D传感器下的动态场景中均有鲁棒性。DynaSLAM在动态场景中的精度优于标准视觉SLAM系统架构，并且可以生成静态的场景地图，而这一点是SLAM在真实环境中长期应用所必需的。

介绍

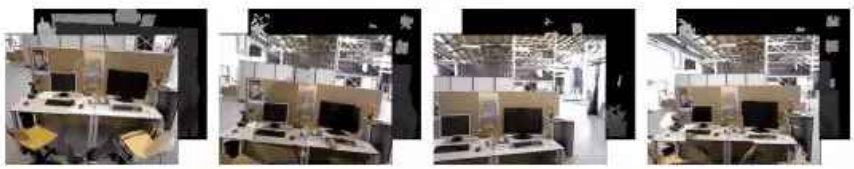
- Dyna-SLAM对动态环境的处理提出了更高的目标：一是尝试去除**所有运动物体**的影响（包括当下保持静止的具有运动能力的物体），二是在建图中抹去所有运动物体，建立包含所有静态物体的三维图像。
- 处理动态环境的思路与先前DS-SLAM基本一致，基于**语义分割(Mask R-CNN)**与**几何上的运动判断(初步预测位姿后的深度估计)**
- Dyna-SLAM将处理分为两种情况，一是mono&stereo，二是RGB-D。主要由于希望通过RGB-D的深度信息进行多视图几何下的运动判断。（stereo也可以计算深度信息，但此处可能由于对深度的精确度要求较高，所以不列入stereo）

Contribution:

1、提出了基于ORB-SLAM2的视觉SLAM系统，通过增加语义分割与多视图方法使得其在单目、双目、RGB-D相机的动态环境中均具有稳健性。

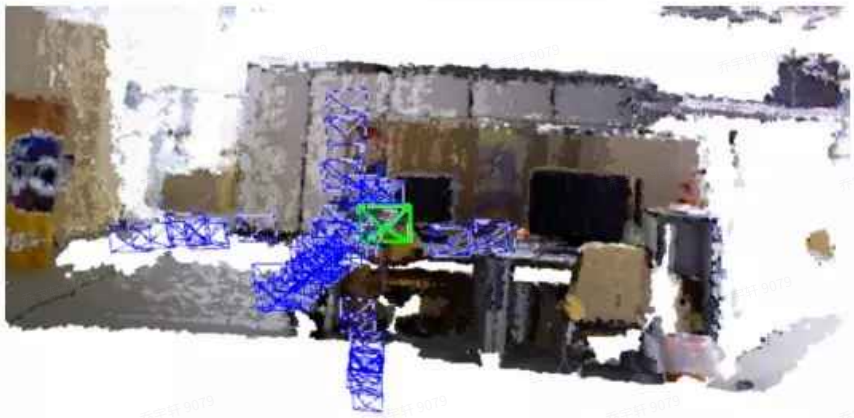
2、通过对因动态物体遮挡而缺失的部分背景进行修复，生成一个静态场景地图。

(a) Input RGB-D frames with dynamic content.



(b) Output RGB-D frames. Dynamic content has been removed. Occluded

background has been reconstructed with information from previous views.



(c) Map of the static part of the scene, after removal of the dynamic objects.

图1 RGB-D例子下DynaSLAM的结果

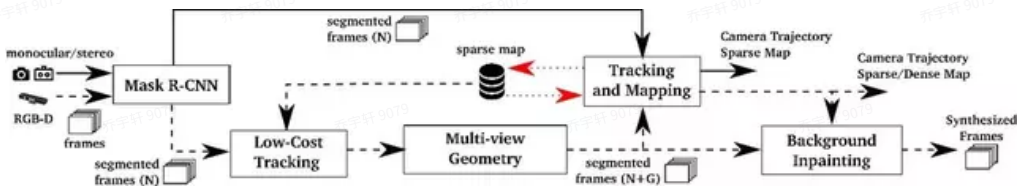


图2 本文方法的完整流程图

对于RGB-D相机而言，将RGB-D数据传入到CNN网络中对有先验动态性质的物体如行人和车辆进行逐像素的分割。作者使用多视几何在两方面提升动态内容的分割效果。首先作者对CNN输出的动态物体的分割结果进行修缮；其次，将在大多数时间中保持静止的、新出现的动态对象进行标注。对于单目和双目相机，则直接将图像传入CNN中进行分割，将具有先验动态信息的物体分割出去，仅使用剩下的图像进行跟踪和建图处理。

内容细节

A、使用卷积神经网络对潜在的动态物体进行分割

作者使用Mask RCNN进行实例分割，他们认为在大多数环境中的潜在动态或可移动物体有：人、自行车、汽车、猫、摩托车、飞机、人，自行车，汽车，摩托车，飞机，巴士，火车，卡车，船，鸟，猫，狗，马，羊，牛，大象，熊，斑马和长颈鹿。如果需要增加其他类别，可以在MS COCO数据集上进行微调得到相应的权重模型。在实例分割部分，输入数据为m*n*3的RGB图像，输出为m*n*L的矩阵，再将L层分类图像合并成一幅图像。

引入Mask R-CNN是为了先验地分割出具有运动能力的物体。即使他们处在静止状态，我们也可以在之后的SLAM过程中剔除这些不稳定因素，实现对动态环境的鲁棒，同时Mask R-CNN也是物体识别分割方法中最先进的方法之一。

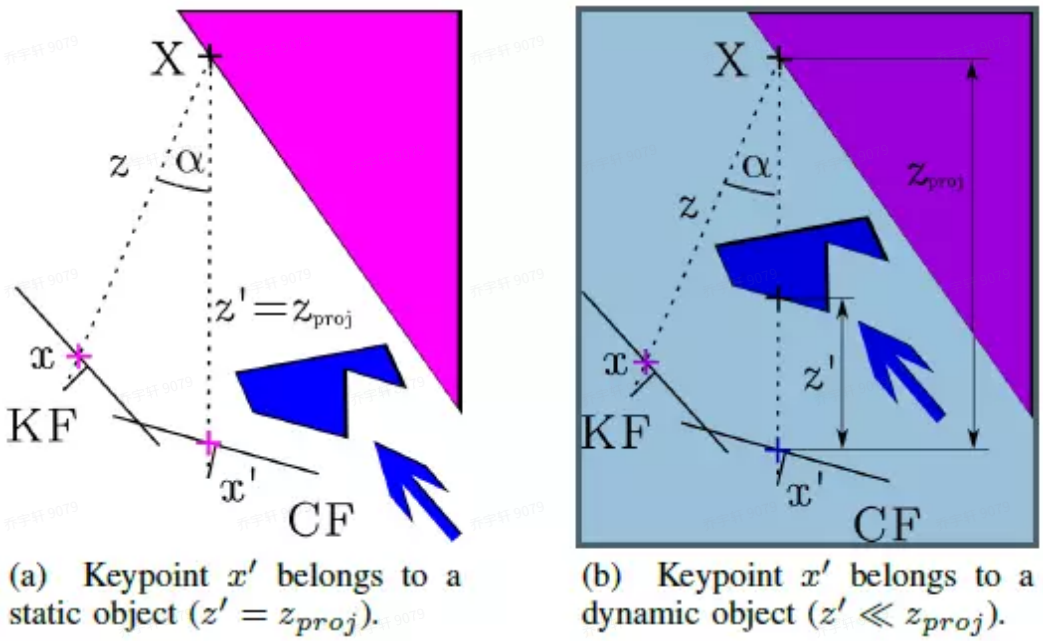
B、低成本追踪

我们希望基于相机的转移估计深度，再次之前需要对相机的运动进行估计。此处我们ORB-SLAM2中追踪相机算法的轻量版，仅仅最小化重投影误差。注意此时动态物体已经被分割，会导致分割曲线附近出现原本不是特征点的新角点，从而为了消除这一影响，我们不考虑所有分割曲线附近的特征点。

C、基于Mask RCNN和多视几何的动态物体分割

主要是针对性地处理在Mask RCNN中没有先验动态标记而具有移动性的物体的分割，例如行人手中的书等。

对于每一个输入影像帧，作者选择一些与其重叠度最大的旧影像帧（文中作者选择数量为5），将这些旧影像帧上的关键点x投影到当前帧上得到特征点x'，以及其投影深度zproj，同时生成对应的三维点X。计算关键点x，x'与三维点X形成的夹角xXx', 记为Alpha，若Alpha大于30度则认为该点可能被挡住了，即不对其做处理。作者观察到在TUM数据集中，夹角Alpha大于30°时的静态物体即被认为是动态的。单目、双目情况下，作者使用深度测量计得到x'对应的深度值z'，在误差允许的范围内，将其与zproj进行比较，超过一定阈值则认为该点x'对应于一个动态的物体。判断过程如图3所示。作者经过在TUM数据集上进行的测试发现深度值差阈值为0.4m时，表达式0.7*Precision +0.3* Recall达到最大。



(a) Keypoint x' belongs to a static object ($z' = z_{proj}$).

(b) Keypoint x' belongs to a dynamic object ($z' \ll z_{proj}$).

图3 采用多视几何判断动态物体示意图

作者对处于运动物体边缘的被标记为动态物体的关键点进行了额外处理，通过其邻域范围内的像素进行判断，对其标记状态进行修改。对图像中像素的分类标记使用深度图上的**区域生长算法**得到了运动物体的掩膜。图4显示了基于多视几何、深度学习和两者结合的方式进行动态物体分割的结果对比

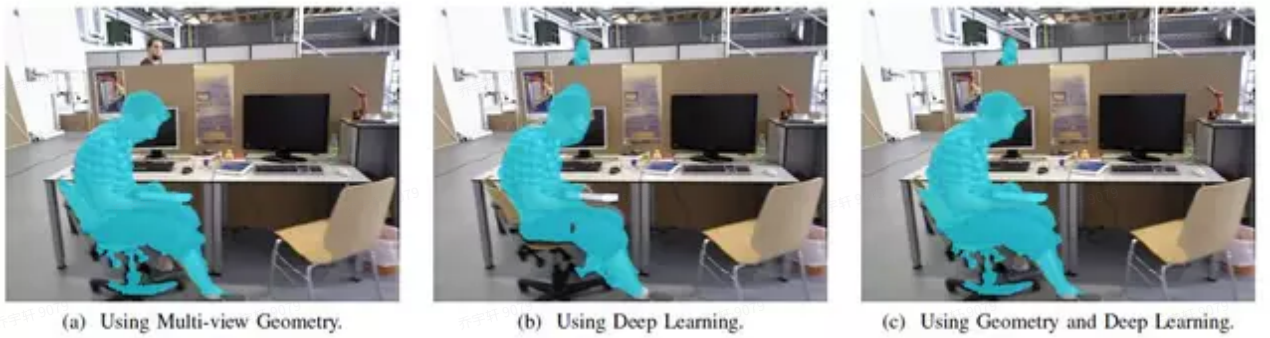


图4 使用不同分割方法检测和分割动态物体的对比结果

我们就图4来说明CNN与多视图几何各自的作用。我们的首要目标是识别所有具有运动能力的物体并对剩余的静态物体建图。对于图4(a)，只采用多视图几何的办法无法识别后方的人；对于图4(b)，CNN无法识别人手中的书，导致建图时会留下一本悬浮的书，这显然不是我们想要的结果。图4(c)展示了结合CNN与多视图几何的结果，实现了对静态物体的建图。

注意到DynaSLAM在由动态特征点生成动态区域时采用的是区域生长算法，即一个动态特征点可能生长为整个区域。当特征点落在边界上时，可能会将静态部分标记为动态。为了避免这一情况的出现，我们利用图像的深度信息。如果动态特征点邻域的深度具有较大的方差，我们将其标签修改为静态。

D、跟踪与建图

我们目前已有了RGB与深度图片，以及分割掩膜，要去掉的特征点包括动态掩膜中的特征点和落在分割曲线附近的特征点。移除后者的原因在于分割曲线周围是高梯度区域。将剩余静态的特征点输入ORB-SLAM2的跟踪与建图线程。

在完成了跟踪与建图后，我们可以利用生成的稀疏地图(即流程图中红色信息交换部分)进一步更新原先的低成本追踪，优化多视图几何的部分。

E、背景修复

作者将之前20关键帧的RGB以及深度图投影到当前帧上完成无动态物体的背景修复。值得注意的是，由于在其他帧中再也没有出现过当前帧中的场景或者深度信息无效造成投影失败，会导致结果中不可修复裂痕的出现。一旦出现这种情况则还需要另寻他法。图5显示了从不同TUM基准序列中得到人工合成影像。利用这些经过修复的影像可以得到满足静态场景假设的SLAM效果。

(c) Inpainted RGB images. (d) Inpainted depth image.

F、关于流程顺序的思考

我们先前所给的流程图中是利用CNN与多视图几何去掉动态特征点以后进入ORB-SLAM2的追踪与建图，记这种流程为N+G。而DynaSLAM中又可以实现背景的修复，即对于一张图片可以去掉其中动态物体并添加上静态背景。如果背景修复的结果足够准确，理论上能带来更多静态特征点。在这种情况下再进行ORB-SLAM2中的追踪与建图也许能够增加准确性，记这样的流程为N+G+BI。

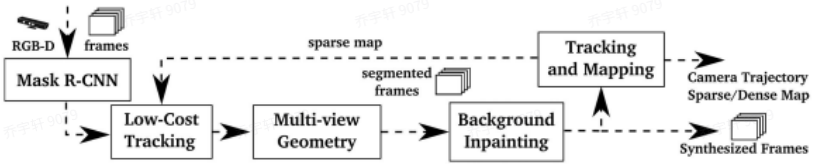


Fig. 6. Block diagram of RGB-D DynaSLAM (N+G+BI).

事实上，N+G+BI不如N+G准确，可以参考两者在TUM数据集上结果的对比。

TABLE I
ABSOLUTE TRAJECTORY RMSE [M] FOR SEVERAL VARIANTS OF DYNASLAM (RGB-D)

Sequence	DynaSLAM (N)	DynaSLAM (G)	DynaSLAM (N+G)	DynaSLAM (N+G+BI)
w_halfsphere	**0.025**	0.035	**0.025**	0.029
w_xyz	**0.015**	0.312	**0.015**	**0.015**
w_rpy	0.040	0.251	**0.035**	0.136
w_static	0.009	0.009	**0.006**	0.007
s_halfsphere	**0.017**	0.018	**0.017**	0.025
s_xyz	0.014	**0.009**	0.015	0.013

可以看到N+G的效果明显优于N+G+BI，添加背景修复的信息反而并没有提高算法的准确率。原因在于，背景修复的过程本身就非常依赖相机位姿的估计。N+G+BI的流程中，背景修复(BI)基于的是先前轻量的低成本追踪，本身准确率就受限，将BI在应用于主体的追踪与建图将误差进一步传递到了ORB-SLAM2的过程中。从而背景修复应当在放在追踪与建图之后，利用更精确的位姿。

问题

- 对动态特征点采用区域生长算法会使仅仅一个动态特征点就生长为动态区域
- Mask R-CNN的速度较慢，DynaSLAM在时间上与DS-SLAM相比没有优势
- 多视图几何中的一些参数，如30度等，是否能良好适应所有数据情况
- 当去除了特征点集中的动态特征点后，如果剩下的静态特征点过少如何处理，在复杂的动态道路环境下，这样的问题更容易出现