InLoc: Indoor Visual Localization with Dense Matching and View Synthesis 笔记

第一作者：Hajime Taira

单位：Tokyo Institute of Technology

作者认为相较于城市定位，室内定位的难点在于：

1. 由于室内场景在几何上规模更小，故而在室内的small change将导致图像上的large change；
2. 大部分的室内场景是textureless and textured areas are typically rather small。这会导致特征匹配经常聚集在图片上的一个小的区域中，造成不可靠的位姿估计；
3. 场景中经常出现对称的重复的元素；
4. 室内场景的在一天中的光照变化相当大；
5. 室内场景经常具有高动态性（人的走动，物体的移动等）。

作者认为目前数据集存在的问题在于：大部分的数据集关注点在于重建、场景理解，而不适合做定位。为了解决这个问题，本文提出一个新的室内定位数据集，相比之前有两点重要的工作：

1. 数据集is large-scale，包含两个大学的建筑；
2. 使用智能手机获得query images，每个参考3D模型之间有一定的时间间隔。这导致query images和参考3D模型经常 包含较大的改变。

**文章结构：**

1. 提出一个全新的适合大规模室内环境下的视觉定位方法；
2. 创建一个全新的适合大规模室内场景的数据集，其中包含较大的视点变化、家具移动、遮挡物或照明变化导致的query images和3D数据库之间存在较大差异；
3. 本文所提出的方法显示出在0.25-0.5 m误差范围内正确定位的查询百分比绝对提高了17-20％。

**详细介绍InLoc dataset for visual localization**

数据集由RGBD图像数据库组成，并由手持式设备拍摄的单独的RGB查询图像集增强所提供的查询图像在3D地图的全局坐标系中使用手动验证的地面真实6DoF相机姿态（参考姿态）进行注释。

1. **Database**

基本的室内RGBD数据集由277张RGBD全景图像组成，每张全景图片由Faro 3D· 扫描仪 拍摄，包含4000万个3D颜色点。每张图的扫面面积在23.5至185.8平方米之间。通过提取标准透视图（60◦FoV），从每个全景图中获得了36个RGBD透视图图像，其中偏航角为30°，俯仰方向为±30°，采样步长总计为10K。数据库包含重大挑战，例如重复的图案（楼梯，柱子），频繁出现的建筑结构（门，窗户），家具位置变化，人们在场景中移动以及无纹理且高度对称的区域（墙壁，地板，走廊，教室） ，开放空间）。

1. **Query Images**

使用智能手机相机（iPhone 7）拍摄了356张照片，这些照片仅分布在两个楼层DUC1和DUC2中。这些查询照片是在一天的不同时间拍摄的，以捕获各种遮挡物和布局（例如人，家具）以及照明变化。

1. **Reference Pose generation**

对于所有查询照片，我们估计6DoF参考相机的姿态为w.r.t. 3D地图。 每个查询摄像机参考姿态的计算如下：

1. 对于每个query，手工的选择一个视觉上最相似的数据库图；
2. 自动匹配query image和选出的数据库图片，使用affine covariant features and nearest neighbor search followed by Lowe’s ratio test；
3. 计算query相机位姿，并视觉验证重投影。通过使用特征匹配获得的query图像点和场景3D点之间的对应关系，通过P3P-RANSAC，然后通过束调整计算查询姿势。通过检查在相应的RGB全景图中检测到的边缘到查询图像中的重投影，以视觉方式评估获得的姿势。
4. 手动将困难查询与选定的数据库映像匹配。手动注释图像像素和3D点之间的5至20个对应关系。
5. 定量和目视检查。对于所有估计的姿势，测量中值重投影误差，计算为重投影的3D数据库点到查询图像中检测到的最近边缘像素的距离，然后去除由于以下原因而导致的粗略误差（距离超过20个像素）的对应关系，如：遮挡。 对于中位数重投影误差小于5像素的查询图像，手动检查查询图像中的重投影边缘，并最终从356个查询图像中接受329个参考姿势。

**具有密集匹配和视图合成功能的室内视觉定位**

首先列出三点室内定位的挑战：

1. 缺乏稀疏local features
2. Large image change
3. Self-similarity

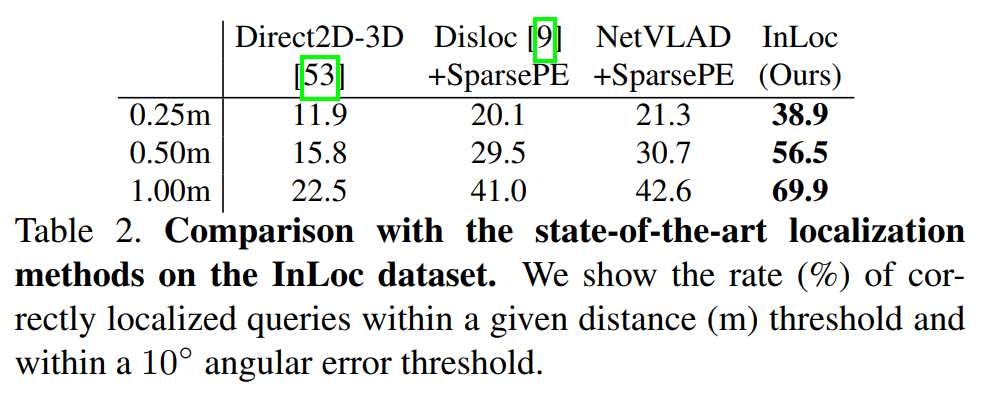
InLoc具有以下三个步骤。 给定查询图像：

1. 通过从注册到地图的参考图像数据库中找到N个最匹配的图像来获得一组候选图像；
2. 对于这N个检索到的候选图像，我们使用与数据库图像一起存储的关联3D信息来计算查询姿势；

使用在规则网格上密集提取的特征，以通过特征匹配和姿态估计来验证和重新排列候选图像。使用了卷积神经网络提取的图像表示形式（VGG-16）。

1. 最后，基于视图综合验证对计算出的相机姿态进行重新排序。

**实验结果**



在InLoc数据集上使用本文方法正确定位的查询比例最高。

数据集：<http://www.ok.sc.e.titech.ac.jp/INLOC/>