

视觉定位与感知课程作业（期末大作业）

通过本课程的学习，我们知道通过若干张图像，我们可以同时获得图像中所拍摄场景的三维信息，并得到图像拍摄时的 6 个自由度的姿态和位置。

本课程大作业内容是使用自己的手机或相机，对同一个场景，分别在三个角度拍摄三张图像，最终获得三张图像拍摄时的姿态和位置，以及图像中特征点的三维坐标信息，并评估和验证所实现算法的正确性和精度性能。

一、实现思路

算法实现的思路是，首先选择其中两张图像，通过特征点匹配基础矩阵的求解得到两个拍摄视角的相对位姿，然后使用三角化得到对应特征点的三维点云。对于第三张图，我们可通过相机位姿估算的方式求解其姿态和位置。最后我们可以采用非线性最小二乘法对稀疏三维点和三个相机 Pose 同时优化，得到最终结果。

具体步骤如下：

1. 标定相机的内参
 - (http://www.vision.caltech.edu/bouquetj/calib_doc/),
 - 或使用 matlab 自带标定工具 (matlab 2018b 以上)
 - 打印一个标定图案
 - 注意标定时需要保证相机拍摄到不同姿态的标定板，并让标定板充满画面。
2. 选择一个合适场景，拍摄三张图像
 - 所选择的场景纹理丰富，并且三维结构突出（不要选择一个纯粹平面的场景，庙门是一个不错的场景）
 - 三个拍摄角度间既要有足够的基线，也要控制相互夹角不大于 30 度，不然会造成后续特征点匹配的困难。
3. 去除畸变的影响
 - 见第 7 课第 53 页 PPT，使用像素映射的方式得到没有畸变的图像。
4. 选取其中两张图进行特征点匹配
 - 建议用 SIFT 和 SURF 特征点和特征描述子进行匹配
 - 为了降低 Outlier 的比例，可利用最佳匹配和第二好匹配的特征描述子距离比值进行筛选（如让比值大于 1:0.7）
5. 使用 RANSAC 和 8 点法或 5 点法去估算基础矩阵或本质矩阵
 - 注意在理论抽样次数基础上再叠加额外迭代次数
 - 迭代结束，将所有 inlier 拿进去再算一遍基础矩阵，以提升估算精度。
6. 从本质矩阵中分解 R, t ，并对多个解进行筛选得到唯一解
 - 结合下一步进行判断合理的 R, t 组合
7. 利用三角化进行三维点云重构
 - 采用齐次坐标和非齐次坐标两种计算方法都可以

8. 将两张图中的一张与第三张进行特征匹配 (类似步骤四), 利用已经三角化的特征点, 得到第三张图的 3D-2D 对应关系
9. 使用 RANSAC 或 M-estimator 实现第三个视角的相机位置姿态估算。
 - RANSAC 过程中可采用使用开源或自己实现的 P3P, PnP, ePnP 或者 POSIT 方法实现 3D-2D 位姿估算, 然后调用非线性最小二乘法迭代更新位姿
10. (可选) 利用非线性最小二乘法, 对所求解出的 N 个三维点坐标 $\{X_1, \dots, X_i, \dots, X_N\}$, 和三个相机的位姿 $\Theta_1, \Theta_2, \Theta_3$ 进行 Bundle Adjustment 优化。

$$\min \sum_{j=1}^3 \sum_i^N \|x_{ij} - \mathcal{P}(X_i, \Theta_j)\|^2 \quad (1)$$

其中 $x_{ij} - \mathcal{P}(X_i, \Theta_j)$ 表示第 i 个三维点在第 j 个视角的重投影误差。

完成前面 9 个步骤视为作业完成。

二、评估验证

- 评估至少需要包含定性分析, 验证算法实现的正确性。
- 亦可通过仿真实验和真实实验加入定量分析, 对所实现算法的关键步骤或整体性能进行量化评估 (可选)。

三、提交要求

压缩包命名格式: 学号-姓名。压缩包内应包括代码、报告、图片 (代码输入)。报告提交 word 原始版本+PDF 版本, 应包括实验原理, 算法思路 (可包含伪代码), 实验结果分析, 文档采取字体小四, 行间距多倍行距, 1.25。程序单独附上环境配置编译运行的说明文档 (建议用 markdown 文件)。

四、评分标准

情况 A. 所有步骤完成且结果正确

- 70 基本分

情况 B. 未完成

- 按步骤完成度给分(3-9 步, 每个步骤 10 分)

加分项（与基本分累积不超过 100）：

- 完全自己动手实现第 3、5、6、7、9 步骤算法（不调用现有软件包）(每个 +5)
- 文档齐全程序编译运行方便；(+2~5)
- 程序界面友好可视化直观 (+1~3)
- 步骤 10 的实现 (+5)
- 量化评估实现 (+2~5)
- 实验全面，评估方法合理，分析论证详细全面；(+5)

扣分项：

- 完成但结果不正确，视 bug 严重程度扣分