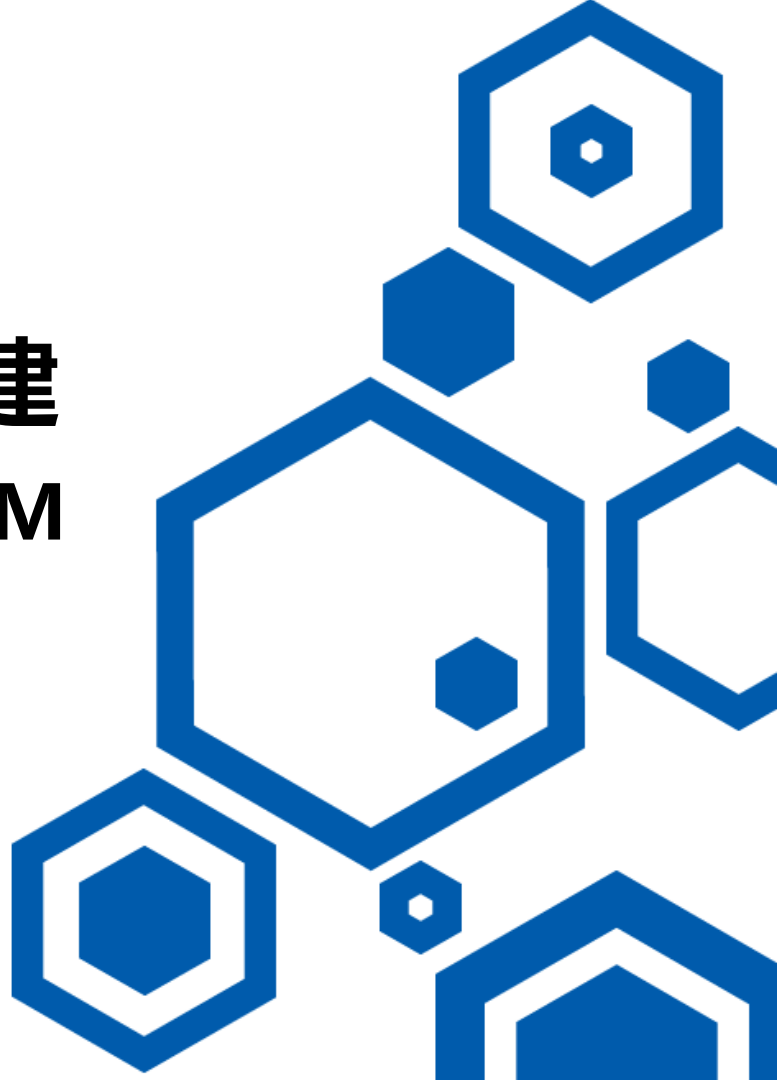


基于图像的三维模型重建

——增量SFM



✓ 几种BA的形式

- ✓ 同时优化相机和三维点
- ✓ 优化相机
- ✓ 只优化三维点
- ✓ 单目相机

✓ 运动恢复结构

- ✓ 增量运动恢复结构 (Incremental SFM)
- ✓ 全局运动恢复结构 (Global SFM)
- ✓ 分层式运动恢复结构 (Hierarchical SFM)

✓ 运动恢复结构的几个问题

数学模型

$\chi_{ij} = 1$ 表示第 i 个点在第 j 个相机中可见;

$\chi_{ij} = 0$ 表示第 i 个点在第 j 个相机中不可见;

$\mathbf{X}_i = (X_i, Y_i, Z_i)^T$ 表示三维点世界坐标系中的坐标;

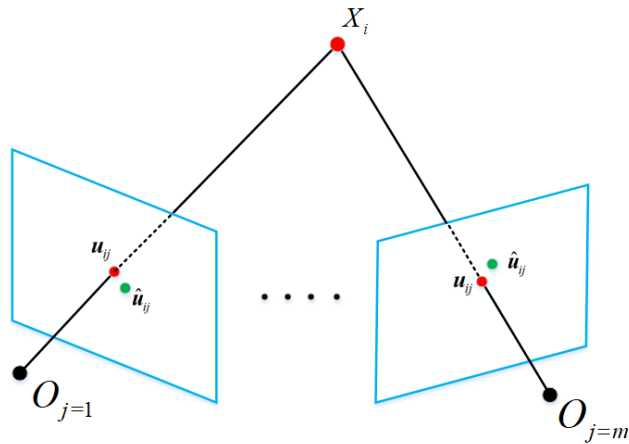
\mathbf{C}_j 表示第 j 个相机的内参数以及世界坐标系与第 j 个相机的外参数;

$\hat{\mathbf{u}}_{ij}$ 表示第 i 个点在第 j 个相机中的观测点;

\mathbf{u}_{ij} 表示第 i 个点在第 j 个相机中的投影点;

$$g(\theta) = \frac{1}{2} \sum_i^n \sum_j^m \chi_{ij} \left\| \mathbf{u}_{ij} - \hat{\mathbf{u}}_{ij}(\mathbf{C}_j, \mathbf{X}_i) \right\|^2 = \frac{1}{2} \sum_i^n \sum_j^m \chi_{ij} e_{ij}^2$$

其中, $\theta = (\mathbf{C}_1, \dots, \mathbf{C}_m, \mathbf{X}_1, \dots, \mathbf{X}_n)$ 为待优化的量, $\theta \in \mathbf{R}^{9m \times 3n}$ 高维空间的非线性优化。



Levenberg-Marquardt法--增量规方程的求解

对称、稀疏

$$\left(J^T(\theta_t)J(\theta_t) + \frac{1}{\lambda}I\right)\delta\theta = J^T(\theta_t)(\mathbf{x} - f(\theta)) \quad f(\theta) = \begin{bmatrix} \vdots \\ \hat{u}_{ij}(C_j, X_i) \\ \vdots \end{bmatrix}$$

$$\begin{aligned} \downarrow \quad & \tilde{J}_{CC} = J_{CC} + \frac{1}{\lambda}I_{CC} \\ & \tilde{J}_{XX} = J_{XX} + \frac{1}{\lambda}I_{XX} \end{aligned}$$

$$\begin{bmatrix} \tilde{J}_{CC} & J_{CX} \\ J_{XC} & \tilde{J}_{XX} \end{bmatrix} \begin{bmatrix} \delta_C \\ \delta_X \end{bmatrix} = \begin{bmatrix} b_C \\ b_X \end{bmatrix}$$

$$\begin{cases} (\tilde{J}_{CC} - J_{CX} \tilde{J}_{XX}^{-1} J_{XC}) \delta_C = b_C - J_{CX} \tilde{J}_{XX}^{-1} b_X \\ \tilde{J}_{XX} \delta_X = b_X - J_{XC} \delta_C \end{cases}$$

线性方程，共轭梯度法求解

固定三维点坐标，优化相机参数

$$\left(J^T(\theta_t)J(\theta_t) + \frac{1}{\lambda}I\right)\delta\theta = J^T(\theta_t)(\mathbf{x} - f(\theta))$$

$$\begin{aligned} \downarrow \\ \tilde{J}_{cc} &= J_{cc} + \frac{1}{\lambda}I_{cc} \\ \tilde{J}_{xx} &= J_{xx} + \frac{1}{\lambda}I_{xx} \end{aligned}$$

$$\begin{bmatrix} \tilde{J}_{cc} & J_{cx} \\ J_{xc} & \tilde{J}_{xx} \end{bmatrix} \begin{bmatrix} \delta_c \\ \delta_x \end{bmatrix} = \begin{bmatrix} b_c \\ b_x \end{bmatrix}$$

由于三维点坐标固定，所以其雅可比矩阵为 $\mathbf{0}$ ，即

$$J_x = \mathbf{0}$$

$$J_{cx} = J_c^T J_x = \mathbf{0}$$

$$\downarrow \\ \tilde{J}_{cc}\delta_c = b_c$$

固定相机参数，优化三维点坐标

$$\left(J^T(\theta_t)J(\theta_t) + \frac{1}{\lambda}I\right)\delta\theta = J^T(\theta_t)(\mathbf{x} - f(\theta))$$

$$\begin{aligned} \downarrow \quad \tilde{J}_{cc} &= J_{cc} + \frac{1}{\lambda}I_{cc} \\ \tilde{J}_{xx} &= J_{xx} + \frac{1}{\lambda}I_{xx} \end{aligned}$$

$$\begin{bmatrix} \tilde{J}_{cc} & J_{cx} \\ J_{xc} & \tilde{J}_{xx} \end{bmatrix} \begin{bmatrix} \delta_c \\ \delta_x \end{bmatrix} = \begin{bmatrix} b_c \\ b_x \end{bmatrix}$$

由于相机参数固定，所以其雅可比矩阵为 $\mathbf{0}$ ，即

$$J_c = \mathbf{0}$$

$$J_{xc} = J_x^T J_c = \mathbf{0}$$

$$\downarrow$$
$$\tilde{J}_{xx}\delta_x = b_x$$

单目相机-内参数共享

$$g(\theta) = \frac{1}{2} \sum_i^n \sum_j^m \chi_{ij} \|\mathbf{u}_{ij} - \hat{\mathbf{u}}_{ij}(f, k_1, k_2, \mathbf{C}_j, \mathbf{X}_i)\|^2 = \frac{1}{2} \sum_i^n \sum_j^m \chi_{ij} e_{ij}$$

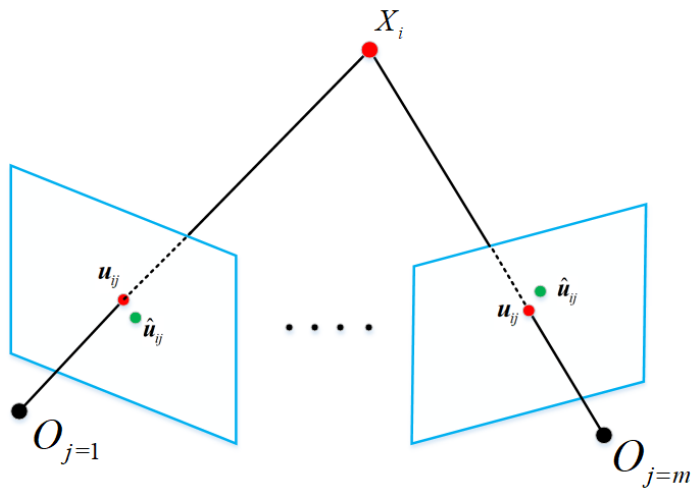
其中, $\theta = (f, k_1, k_2, \mathbf{C}_1, \dots, \mathbf{C}_m, \mathbf{X}_1, \dots, \mathbf{X}_n)$ 为待优化的量,

$\theta \in \mathbf{R}^{3+6m+3n}$ 高维空间的非线性优化。

$$\frac{\partial \hat{\mathbf{u}}_{ij}(f, k_1, k_2, \mathbf{C}_j, \mathbf{X}_i)}{\partial f}$$

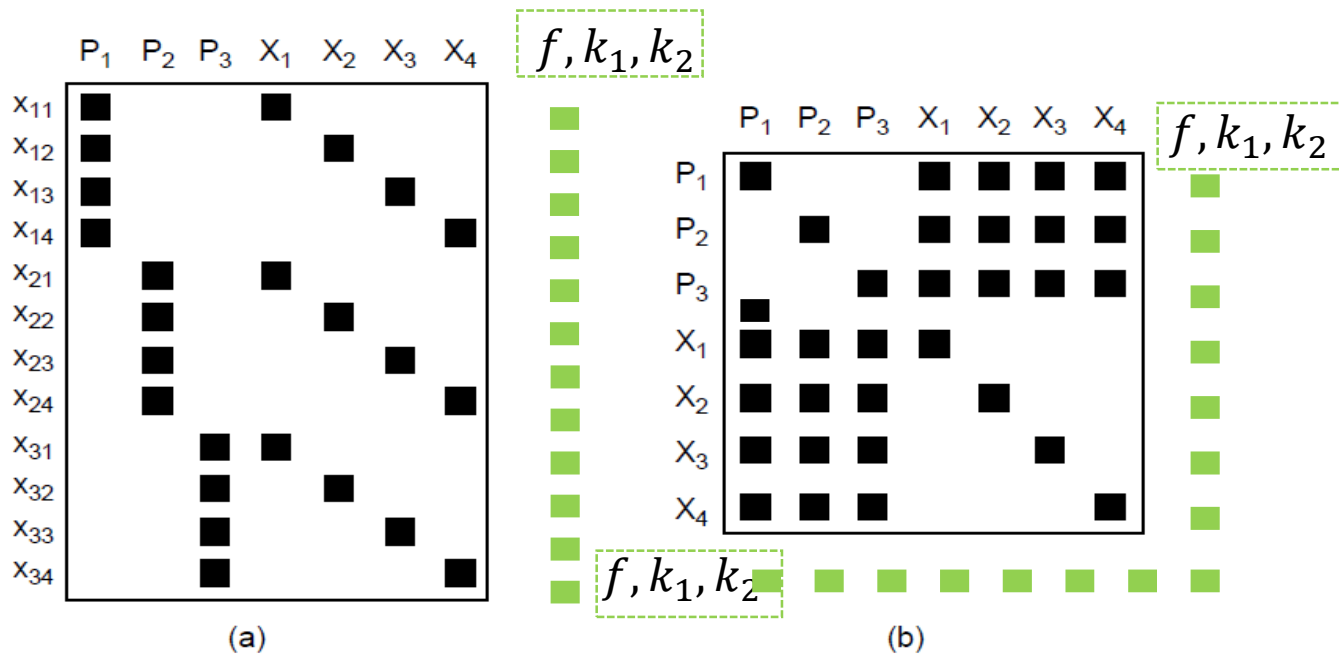
$$\frac{\partial \hat{\mathbf{u}}_{ij}(f, k_1, k_2, \mathbf{C}_j, \mathbf{X}_i)}{\partial k_2}$$

$$\frac{\partial \hat{\mathbf{u}}_{ij}(f, k_1, k_2, \mathbf{C}_j, \mathbf{X}_i)}{\partial k_1}$$



单目相机-内参数共享

相机内参数出现在所有的残差中



✓ 几种BA的形式

- ✓ 同时优化相机和三维点
- ✓ 优化相机
- ✓ 只优化三维点
- ✓ 单目相机

✓ 运动恢复结构

- ✓ 增量运动恢复结构 (Incremental SFM)
- ✓ 全局运动恢复结构 (Global SFM)
- ✓ 分层式运动恢复结构 (Hierarchical SFM)

✓ 运动恢复结构的几个问题

增量运动恢复结构

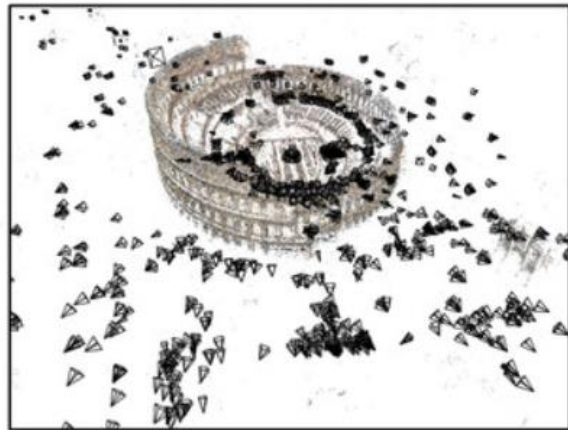
通过相机运动同时恢复相机参数和场景结构

必须有平移 $E = [t]_{\times} R$



无序图像

SFM



稀疏点云

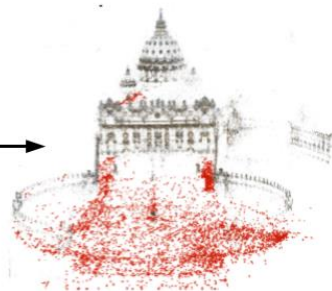
增量运动恢复结构-关键步骤



特征提取与匹配

图像连接图构建

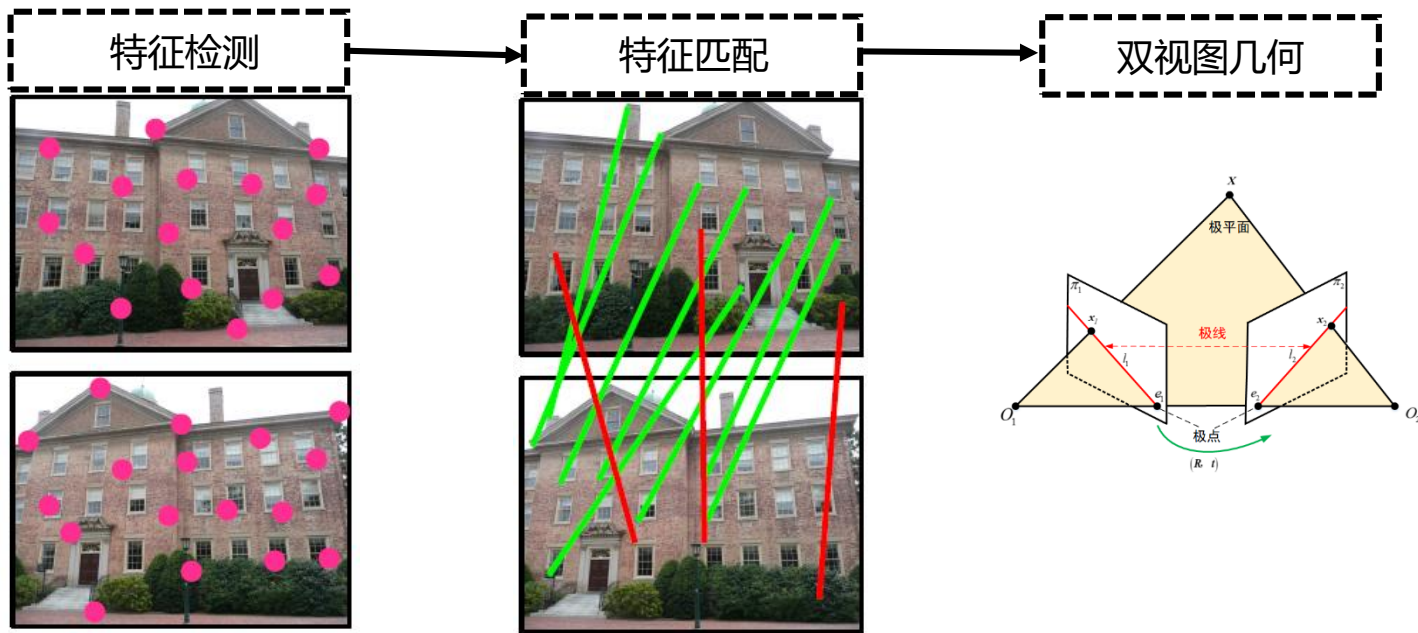
捆绑调整优化相机姿态和场景结构



图像连接图

将两两相对关系对齐到统一的坐标系中

增量运动恢复结构-特征检测与特征匹配



Pairwise matching

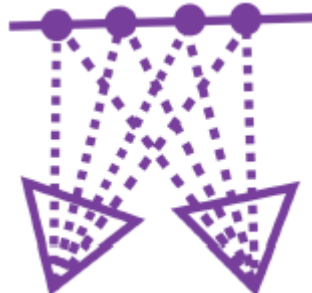
基础矩阵/本质矩阵



基础矩阵(未标定) 7对匹配点

本质矩阵(标定) 5对匹配点

Homography (平面)



场景是平面 4对匹配点

Homography (纯旋转)

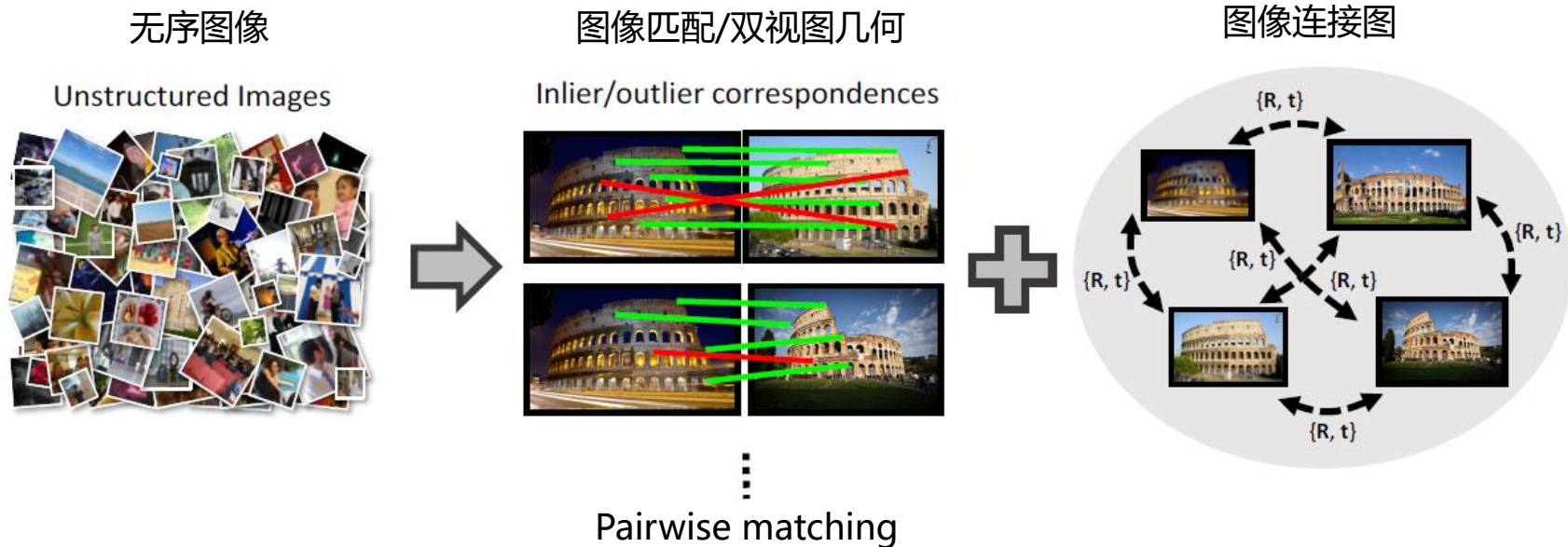


全景图拼接 4对匹配点

增量运动恢复结构-图像连接图

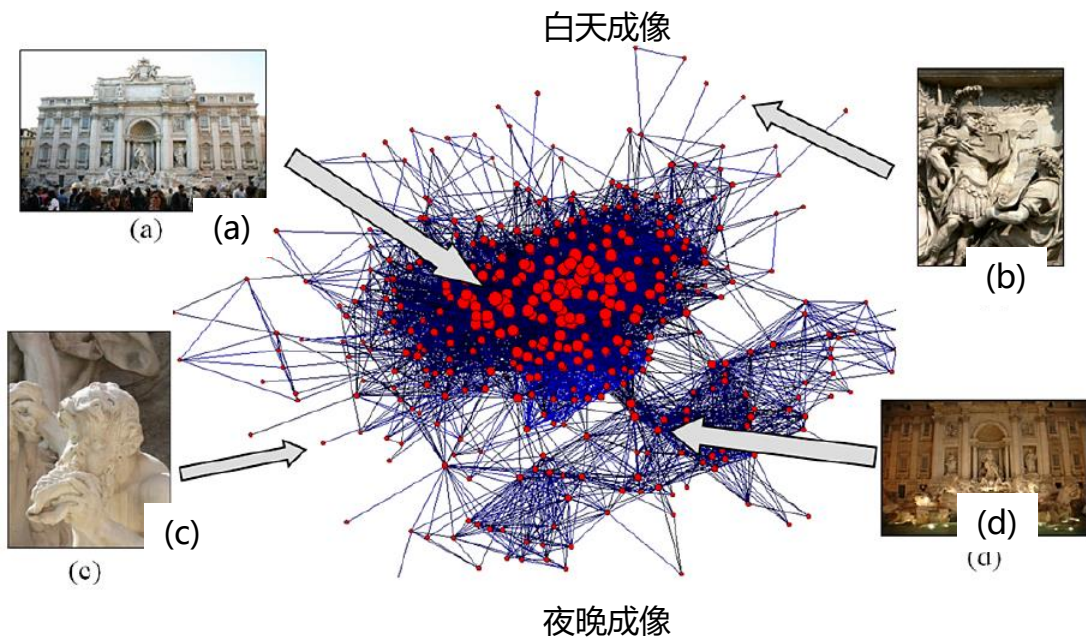
顶点：图像

边界：两幅图像之间存在共同的可见区域（足够多的匹配内点）



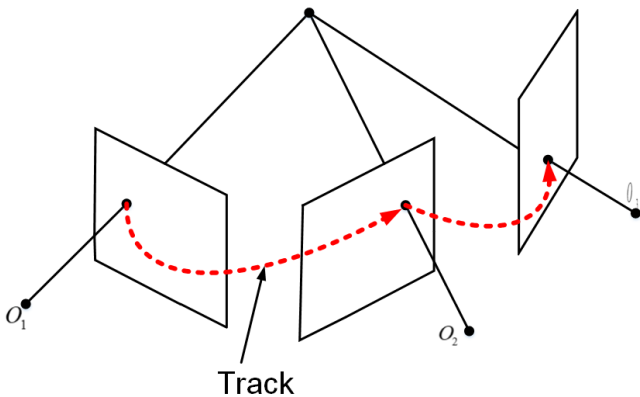
增量运动恢复结构-图像连接图

- ✓顶点的大小与连接个数有关;
- ✓相似的、正面光照条件好的视角集中在中央, 末梢是一些比较偏的视角。



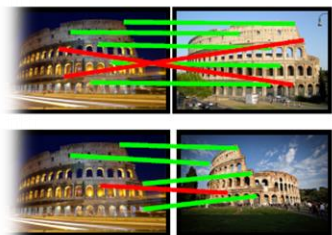
增量运动恢复结构- Tracks

- ✓多个视角对应的匹配点连接起来;
- ✓每一条Track对应一个三维点。

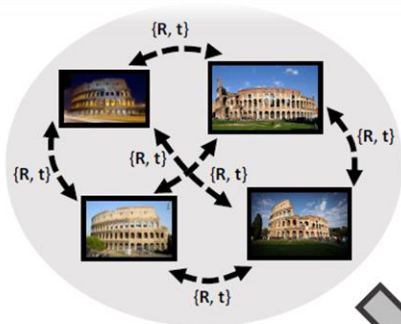


计算Tracks (三维点+匹配关系)

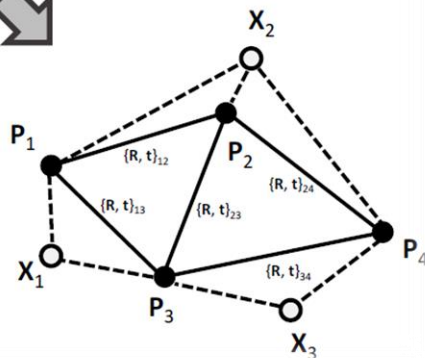
inlier/outlier correspondences



⋮

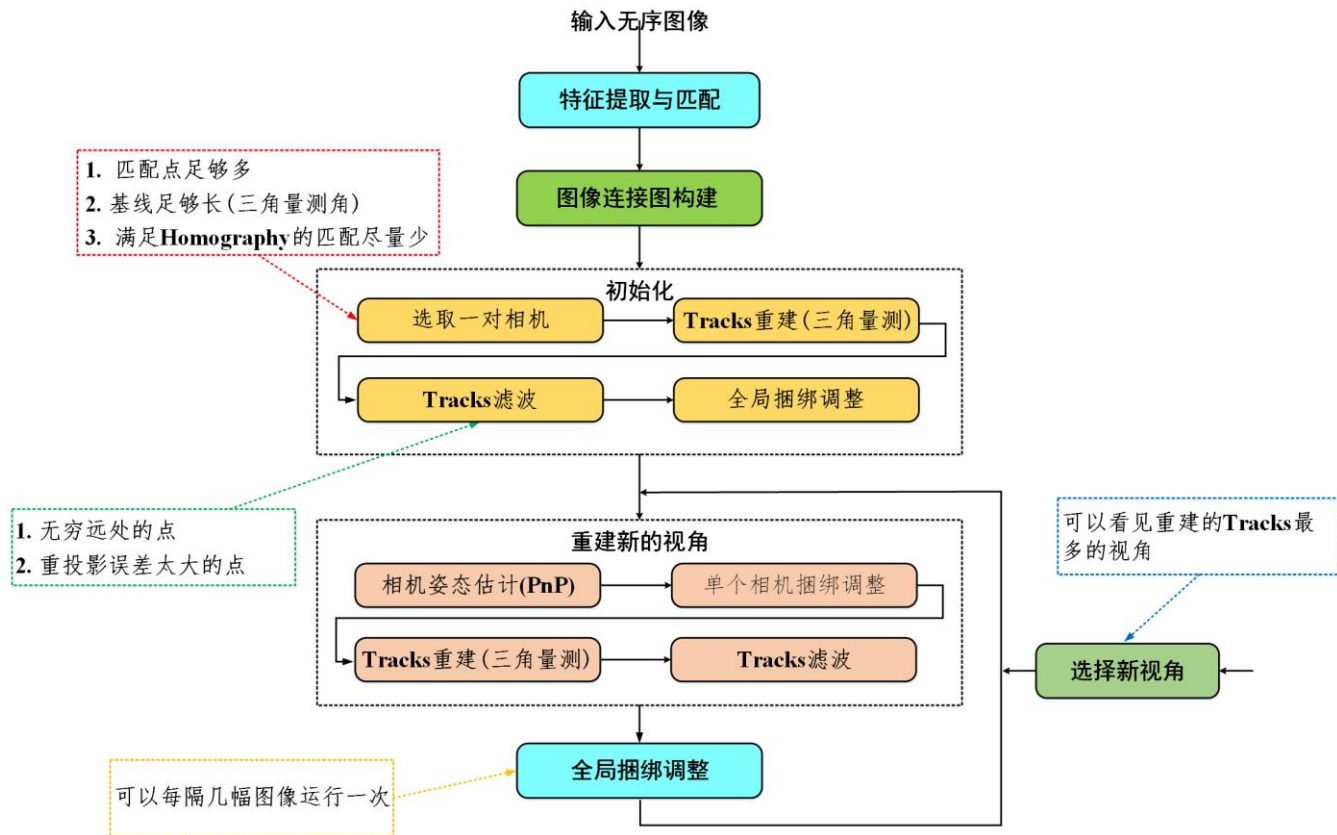


对极几何生成初始相对相机参数



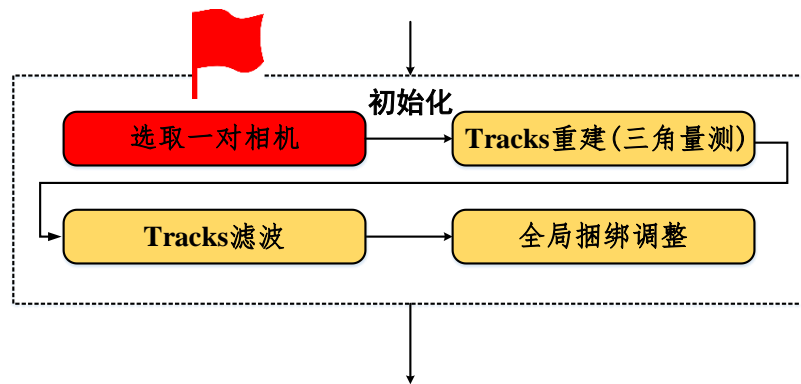
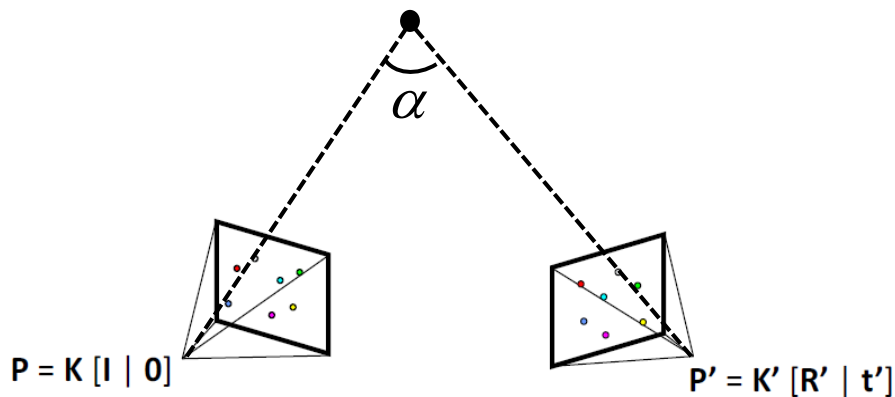
基于RANSAC的三角量测
生成初始三维点

增量运动恢复结构-算法流程



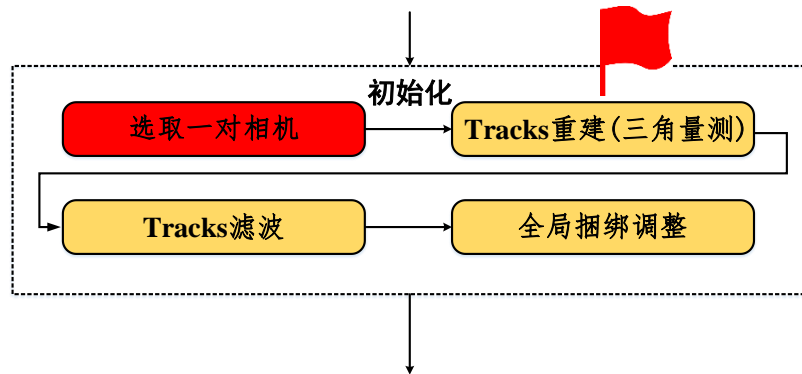
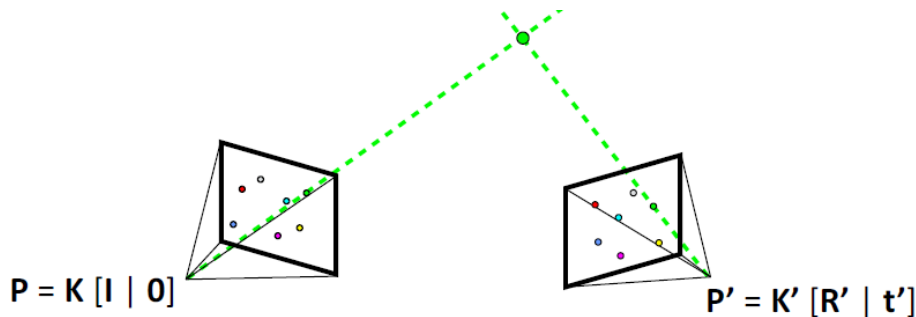
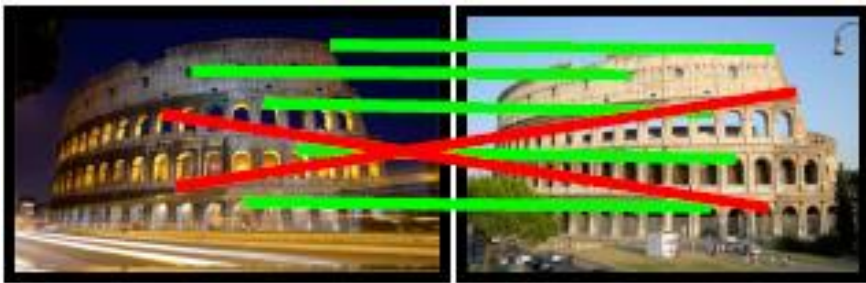
增量运动恢复结构-初始相机对的选取

- ✓ 匹配点足够多(> 50);
- ✓ 基线足够长(三角量测角足够大>5°);
- ✓ 满足Homography的匹配尽量少(内点数<60%);
- ✓ 成功三角化的匹配对>50%。



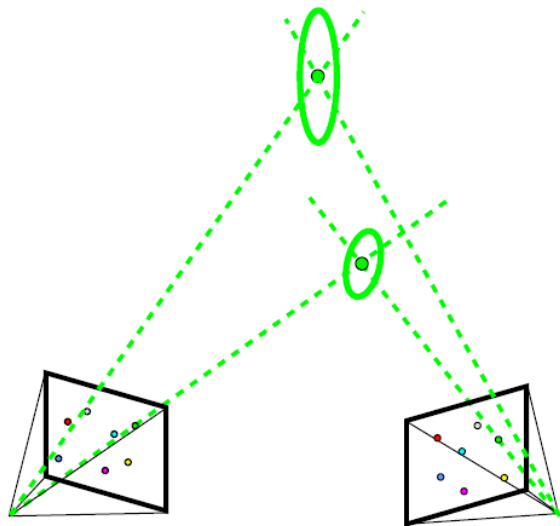
增量运动恢复结构- RANSAC 三角量测

- ✓三角量测将tracks重建成三维点;
- ✓由于存在匹配外点, 重建出错误的tracks, 会严重影响性能, 需要进行tracks滤波

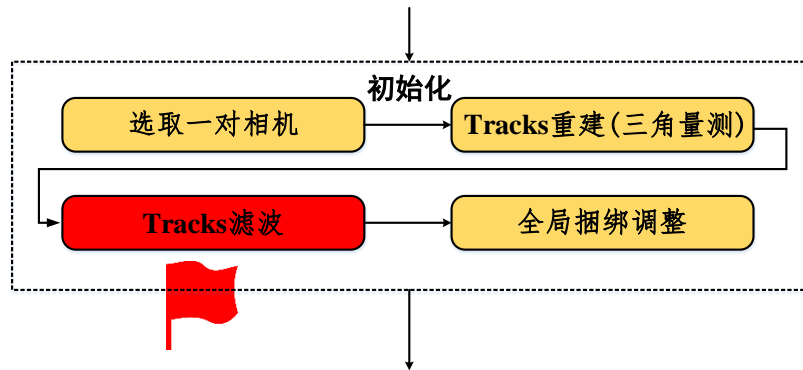
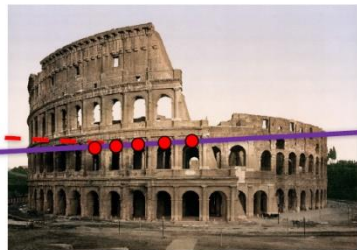


增量运动恢复结构- Tracks滤波

✓ 三维空间中太远的点

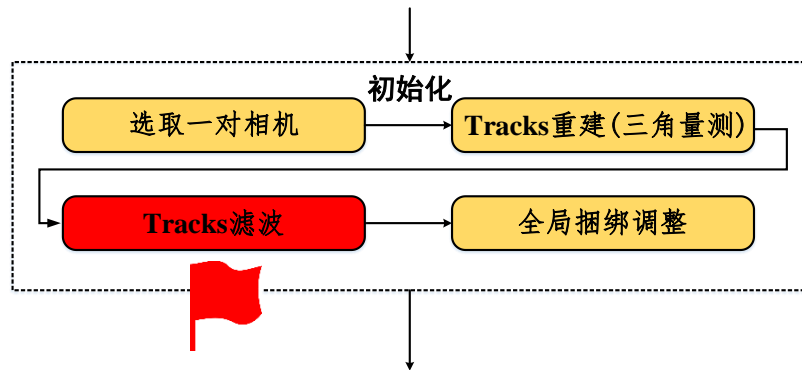
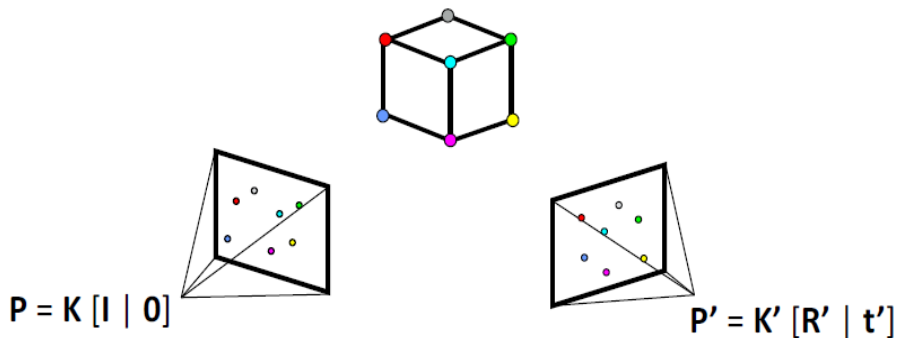


✓ 重投影误差比较大点



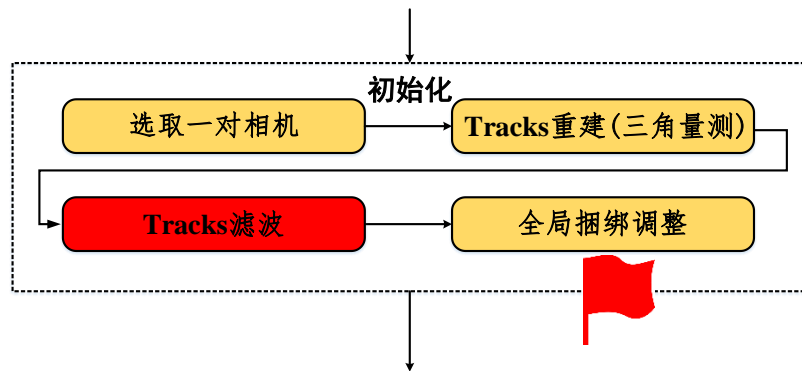
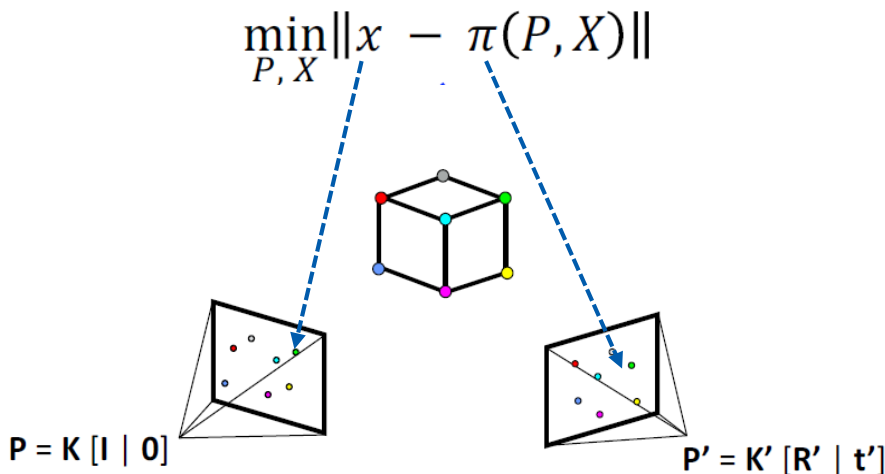
增量运动恢复结构- Tracks滤波

Tracks滤波后的结果



增量运动恢复结构-全局的捆绑调整

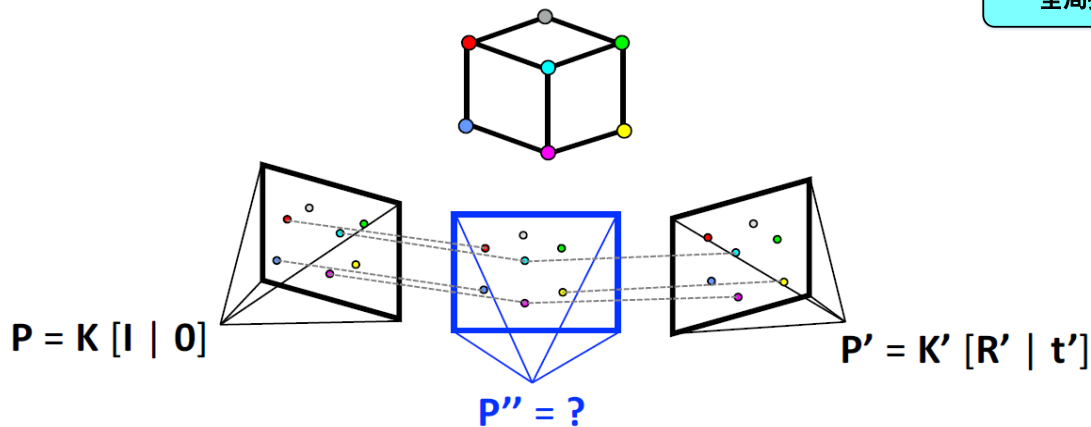
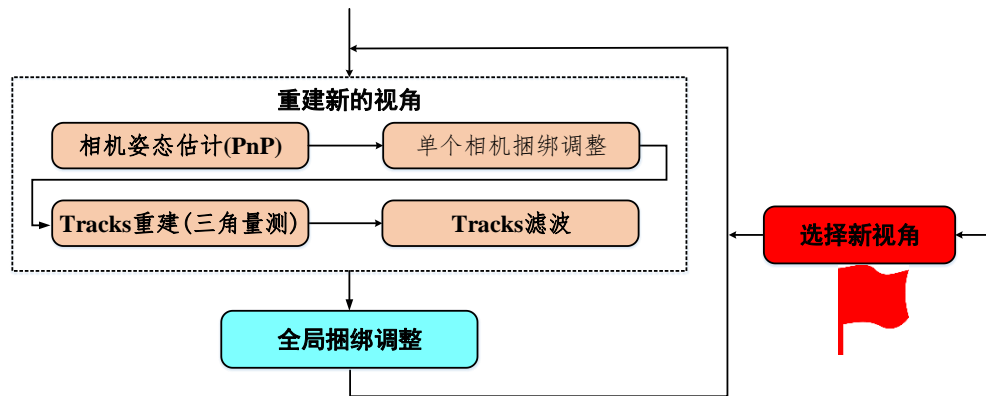
- ✓运动恢复结构的非线性优化部分
- ✓最小化重投影误差



Ceres-Solver, <http://ceres-solver.org/>
Triggs et al., "Bundle Adjustment – A Modern Synthesis"

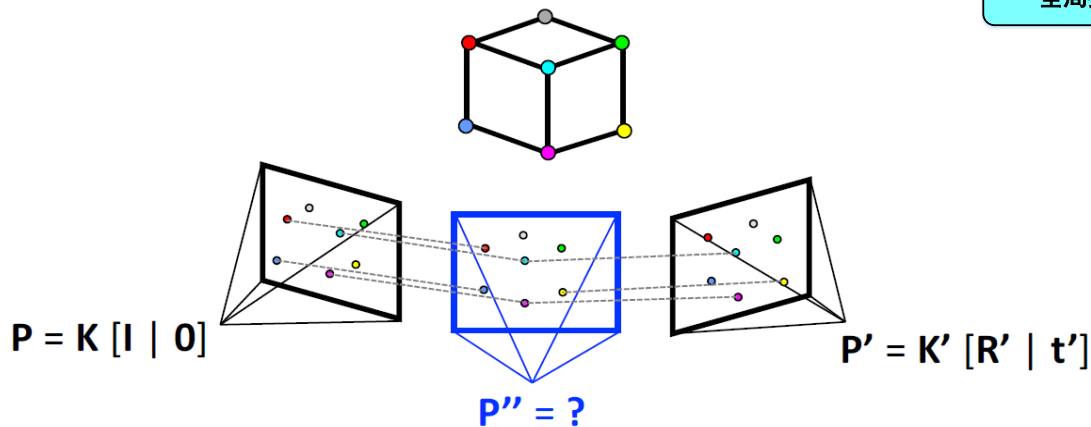
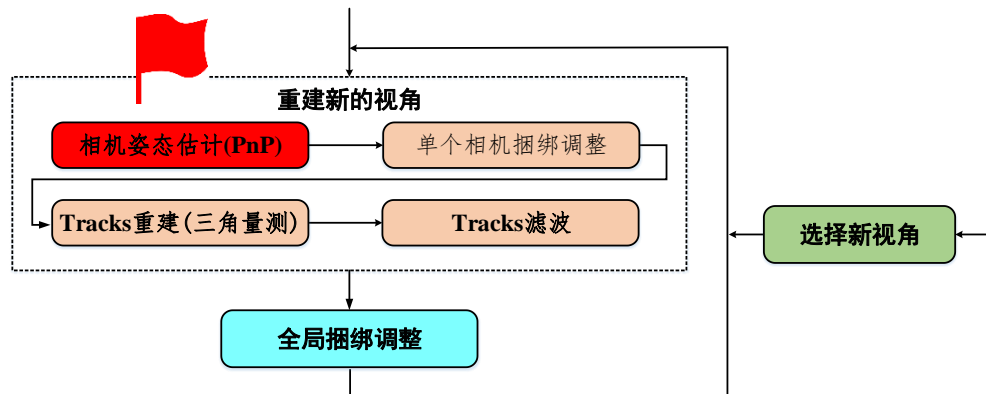
1. 选择新的视角

条件：可见的tracks数目最多



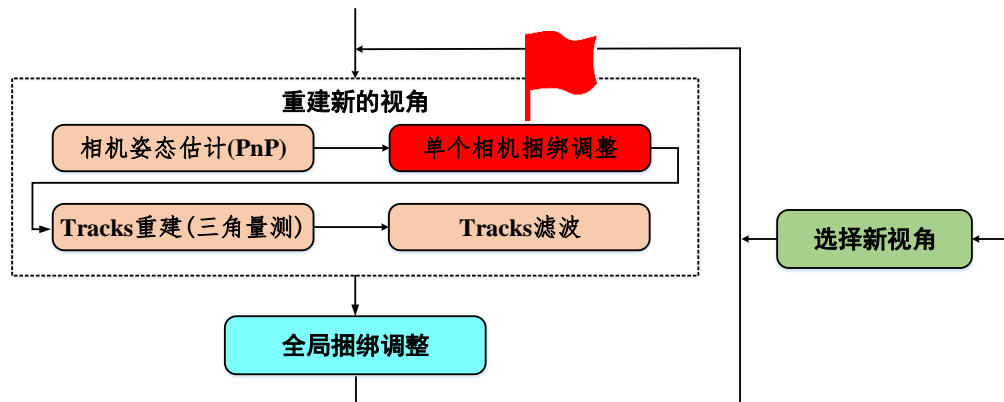
2. 恢复新建相机姿态

- 找到3D-2D对应关系
- 用RANSAC-PnP算法求解相机姿态

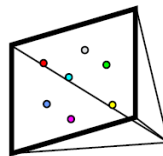
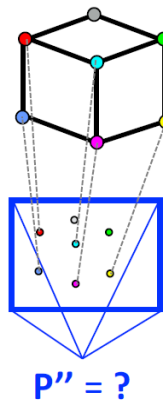
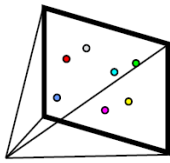


3. 单个相机的捆绑调整

- 相机姿态的非线性优化



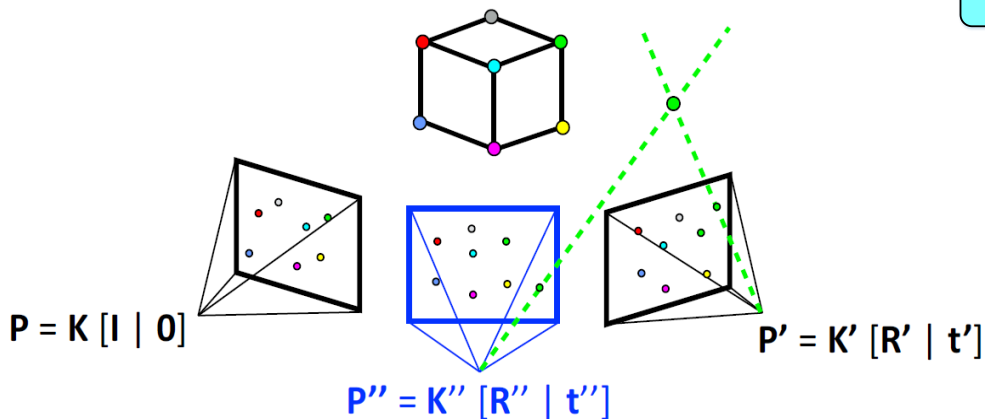
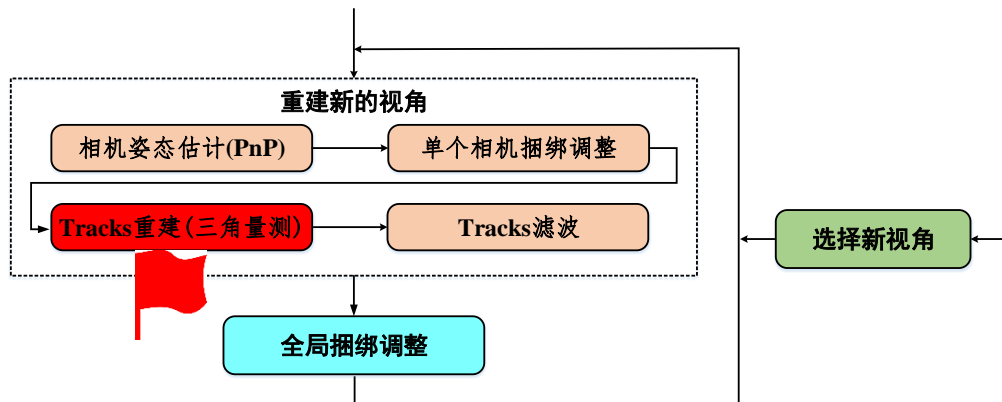
$$P = K [I \mid 0]$$



$$P' = K' [R' \mid t']$$

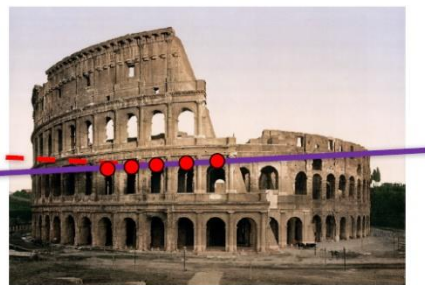
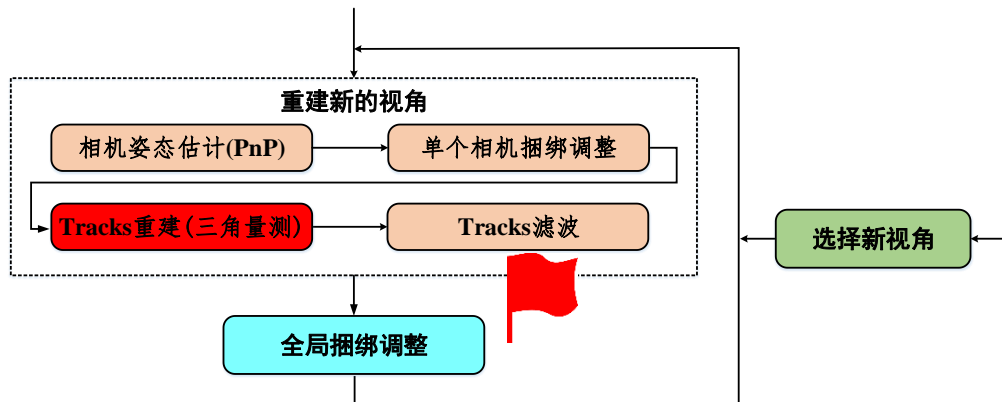
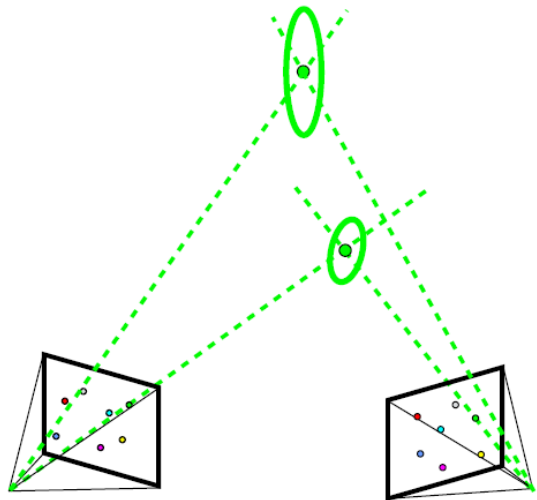
4. Tracks重建

- 增加视角之后, 会产生新的tracks (大于两个视角的匹配点)



5. Tracks滤波

- 对重建的tracks进行外点去除



增量运动恢复结构-全局的捆绑调整

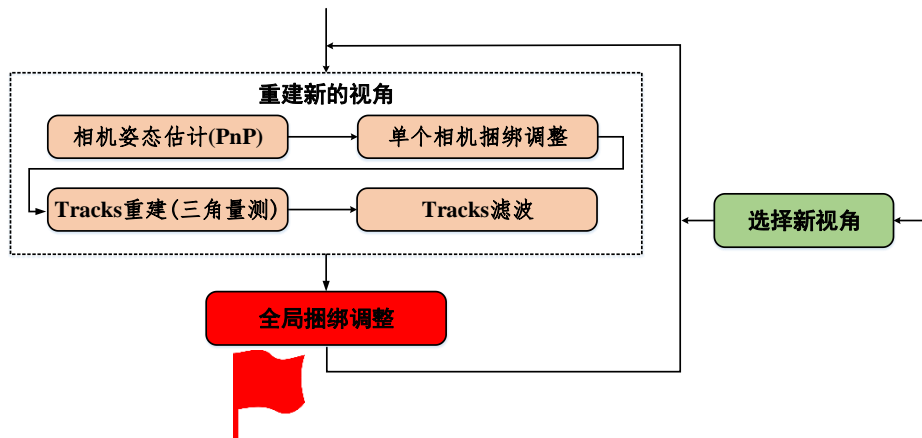
- 对场景中所有的相机和三维点进行非线性优化

$$g(\theta) = \frac{1}{2} \sum_i^n \sum_j^m \chi_{ij} \left\| \mathbf{u}_{ij} - \hat{\mathbf{u}}_{ij}(\mathbf{C}_j, \mathbf{X}_i) \right\|^2 = \frac{1}{2} \sum_i^n \sum_j^m \chi_{ij} e_{ij},$$

$$\theta = (\mathbf{C}_1, \dots, \mathbf{C}_m, \mathbf{X}_1, \dots, \mathbf{X}_n)$$

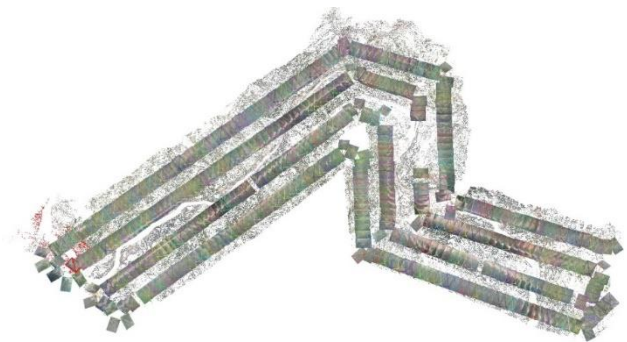
$$\mathbf{C}_j = (f_j, k_{1j}, k_{2j}, \mathbf{R}j, \mathbf{t}_j)$$

$$\mathbf{X}_i = (X_i, Y_i, Z_i)^T$$

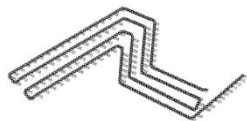


- 每重建一个视角运行一次，也可以重建多个视角运行一次

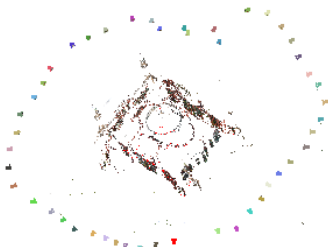
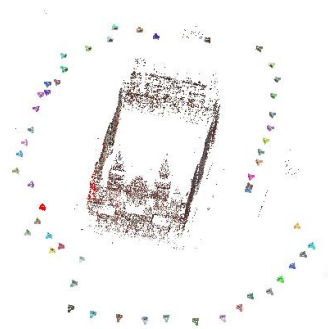
增量运动恢复结构-重建结果



航拍图像稀疏重建结果



航带信息



地面建筑稀疏重建结果

Bundler: <http://www.cs.cornell.edu/~snave/bundler/>

VisualSFM: <http://ccwu.me/vsfm/>

Theia: <http://www.theia-sfm.org/sfm.html>

Colmap: <https://demuc.de/colmap/>

主要优点：

- ✓ 对特征匹配外点鲁棒
- ✓ 重建精度高
- ✓ 捆绑调整不断优化场景结构

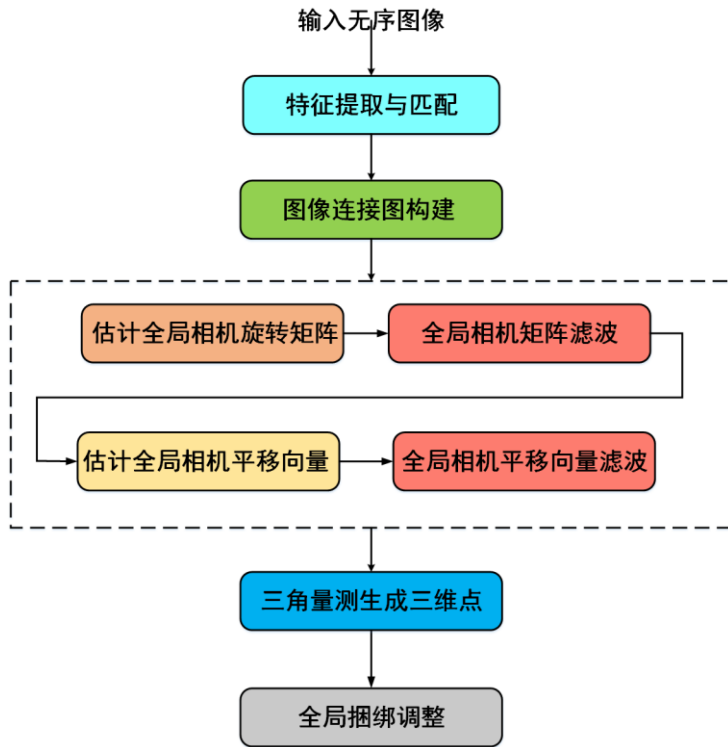
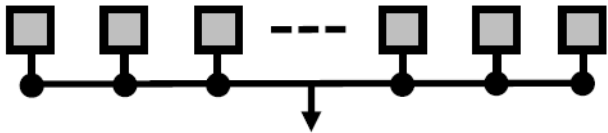
常用来做算法比较的基准

主要缺点：

- ✓ 对初始相机对的选取以及相机添加顺序敏感
- ✓ 大场景产生累计误差导致场景漂移
- ✓ 重复进行捆绑调整，效率低

全局的运动恢复结构

✓ 首先得到所有相机的内外参数,
之后只进行一次全局捆绑调整



全局运动恢复结构

图像连接图

顶点：图像

边界：帧间运动

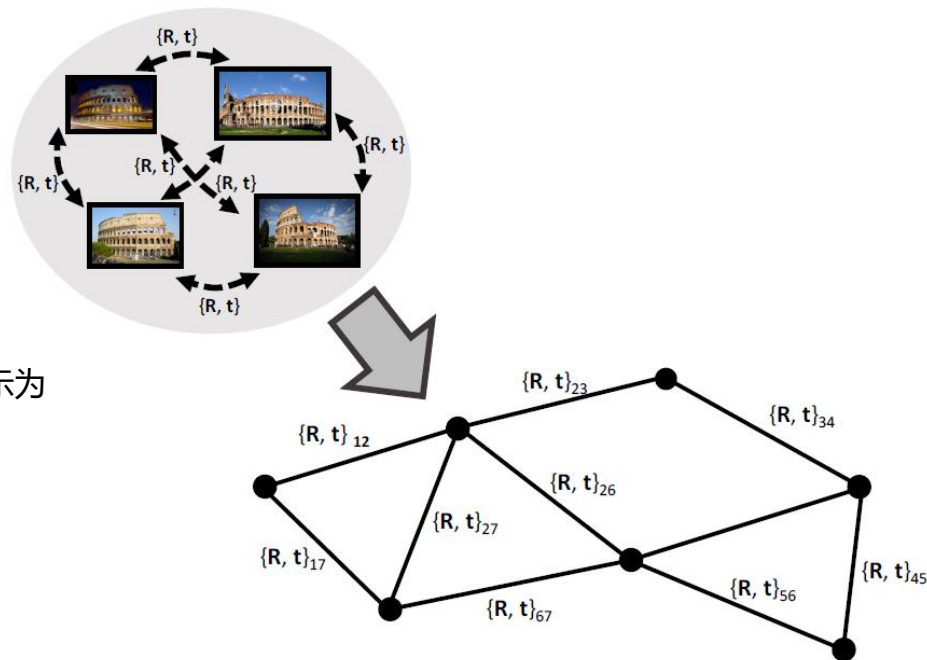
R_i, t_i 表示世界坐标到第 i 个相机坐标的变换

$C_i = -R_i^T t_i$ 表示第 i 个相机在世界坐标系中的位置

R_{ij}, t_{ij} 表示第 i 个相机到第 j 个相机的变换，可以表示为

$$R_{ij} = R_j R_i^T$$

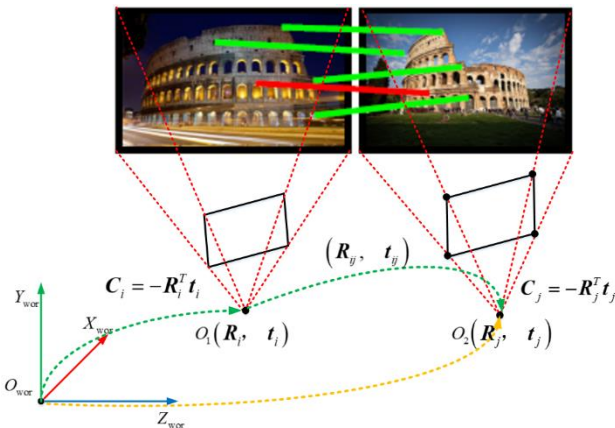
$$\lambda_{ij} t_{ij} = R_j (C_i - C_j)$$



相对约束的推导

$$R_{ij} = R_j R_i^T$$

$$\lambda_{ij} t_{ij} = R_j (C_i - C_j)$$



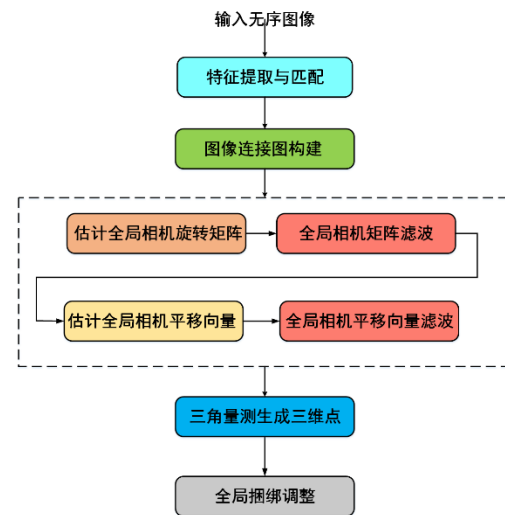
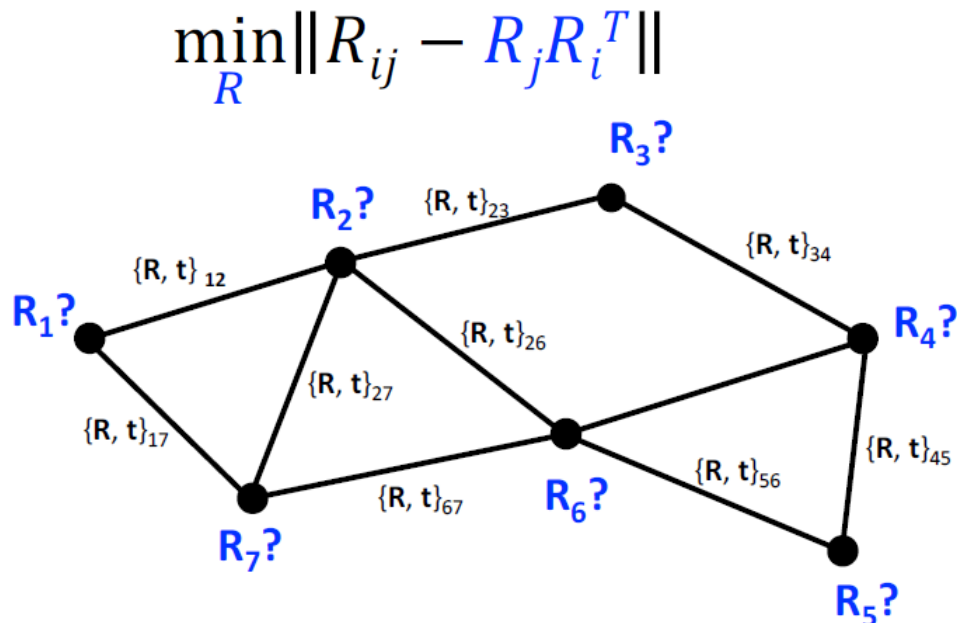
$$R_i^T R_{ij}^T = R_j^T \Rightarrow R_{ij} = R_j R_i^T$$

$$\lambda_{ij} R_i^T (-R_{ij}^T t_{ij}) - R_i^T t_i = -R_j^T t_j$$

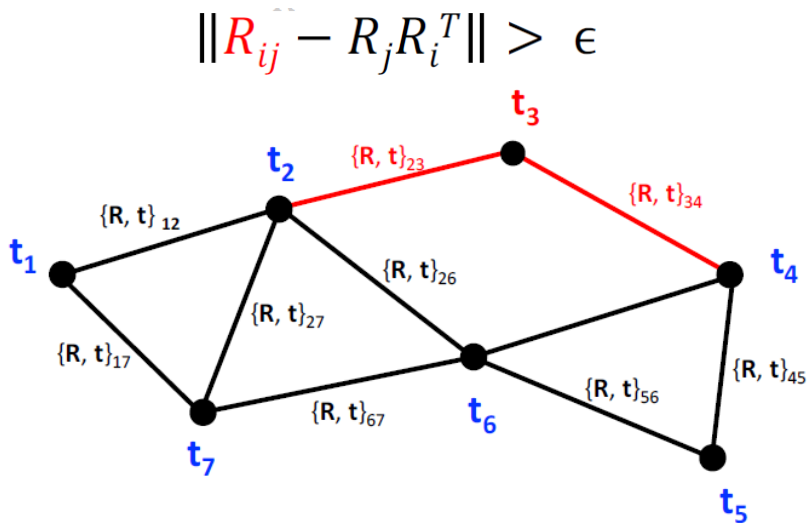
$$\lambda_{ij} t_{ij} = R_j (C_i - C_j)$$

全局的运动恢复结构

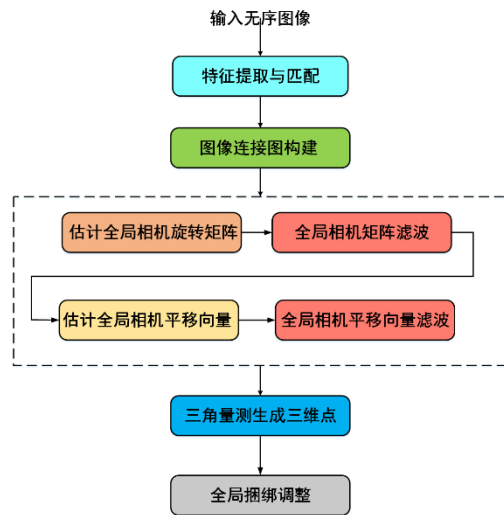
估计全局旋转矩阵-rotation averaging



全局相机矩阵滤波



根据相对旋转矩阵，删除错误的连接边界

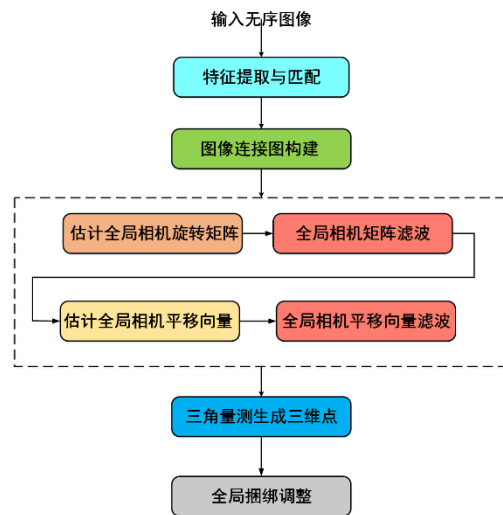
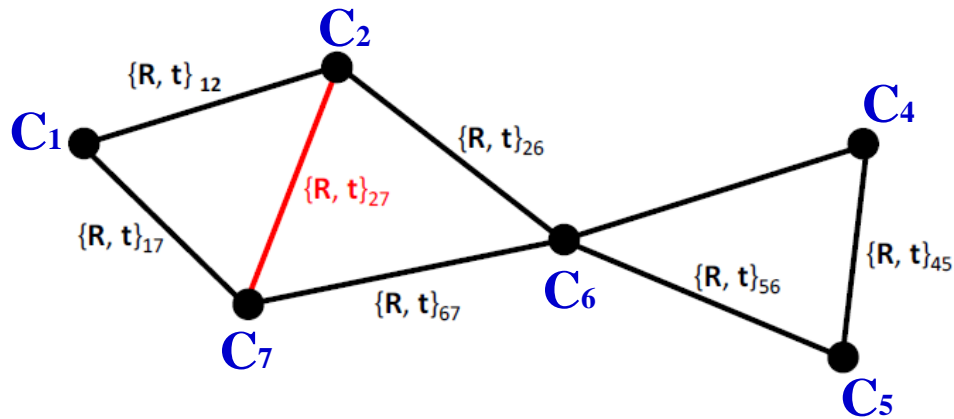


全局运动恢复结构

估计全局平移向量-position averaging

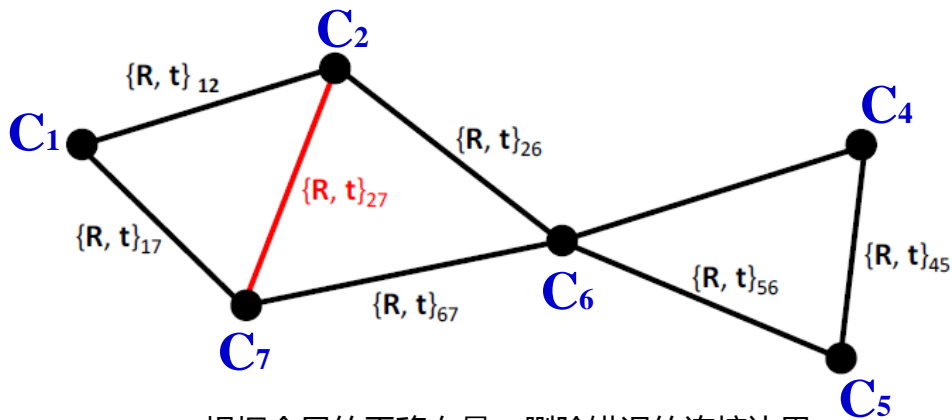
$$\lambda_{ij} \mathbf{t}_{ij} = \mathbf{R}_j (\mathbf{C}_i - \mathbf{C}_j) \quad \longrightarrow \quad \min_c \left\| \mathbf{l}_{ij} - \frac{\mathbf{C}_i - \mathbf{C}_j}{\|\mathbf{C}_i - \mathbf{C}_j\|} \right\|$$

$$\mathbf{l}_{ij} = \mathbf{R}_j^T \mathbf{t}_{ij}$$

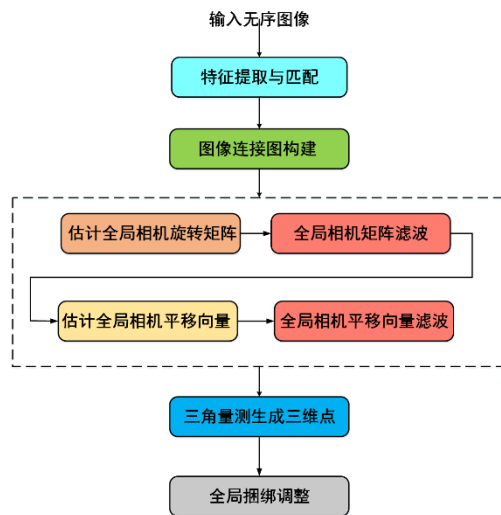


全局平移向量滤波

$$\left\| l_{ij} - \frac{C_i - C_j}{\|C_i - C_j\|} \right\| > \varepsilon$$



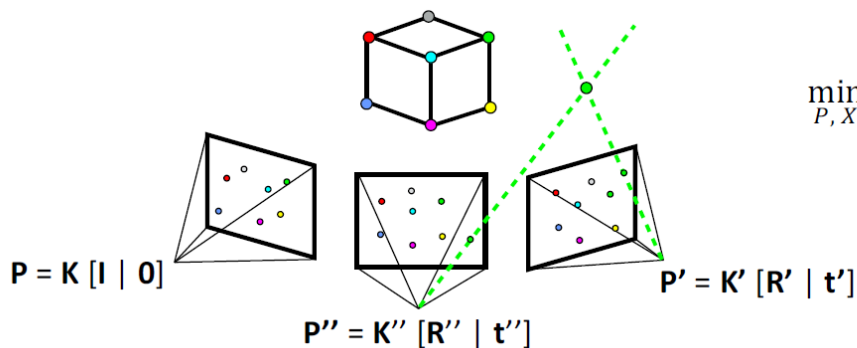
根据全局的平移向量，删除错误的连接边界



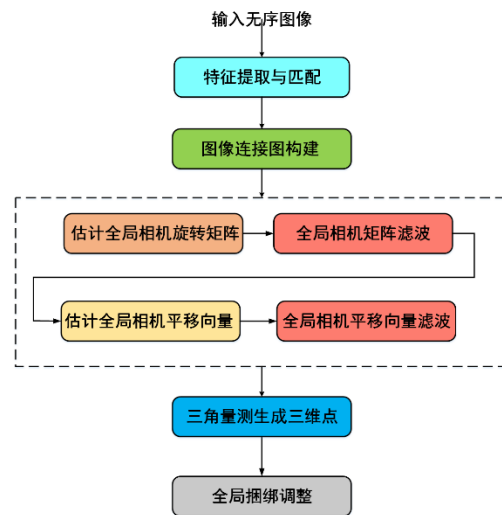
思考：

上述过程和Pose graph有什么区别？为什么要旋转和平移分步进行计算？

三角量测重建三维点并进行全局的BA



$$\min_{P, X} \|x - \pi(P, X)\|$$



全局捆绑调整的优缺点

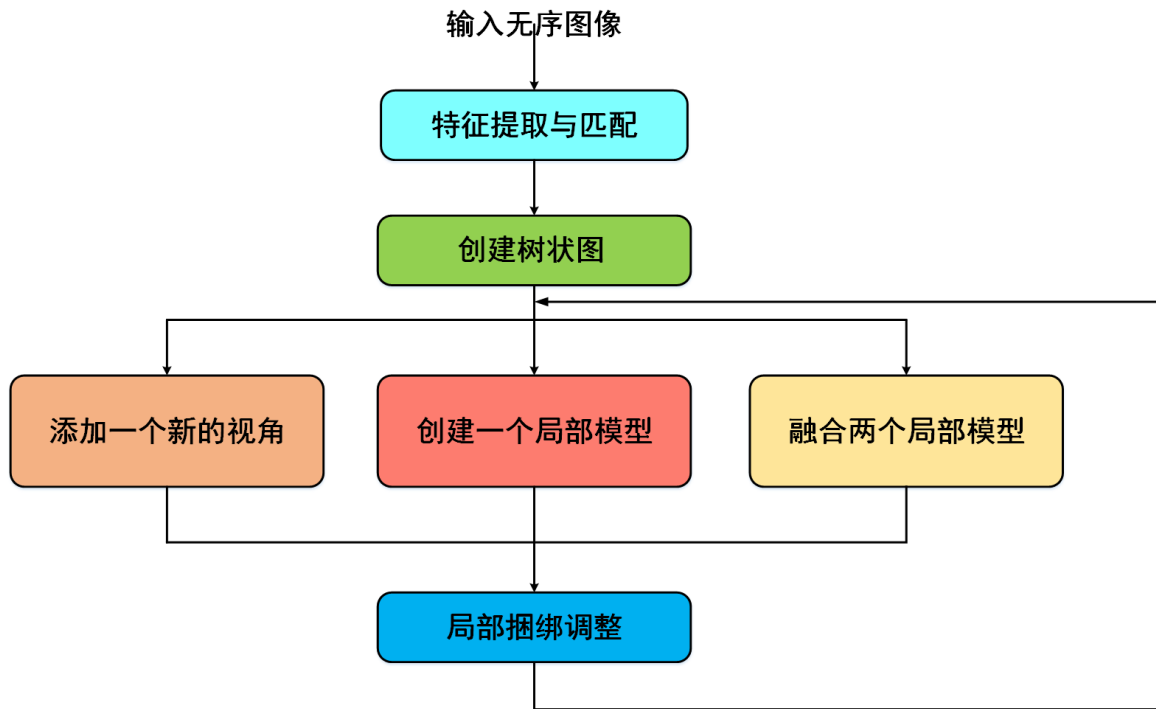
主要优点:

- ✓ 将误差均匀分布在连接图上, 没有误差积累
- ✓ 不需要考虑初始相机选取和相机添加顺序问题
- ✓ 仅执行一次捆绑调整, 重建效率高

主要缺点:

- ✓ 鲁棒性不足, 相机位置求解时对匹配外点敏感
- ✓ 过滤连接图边界, 容易造成部分图像丢失, 导致重建不完整

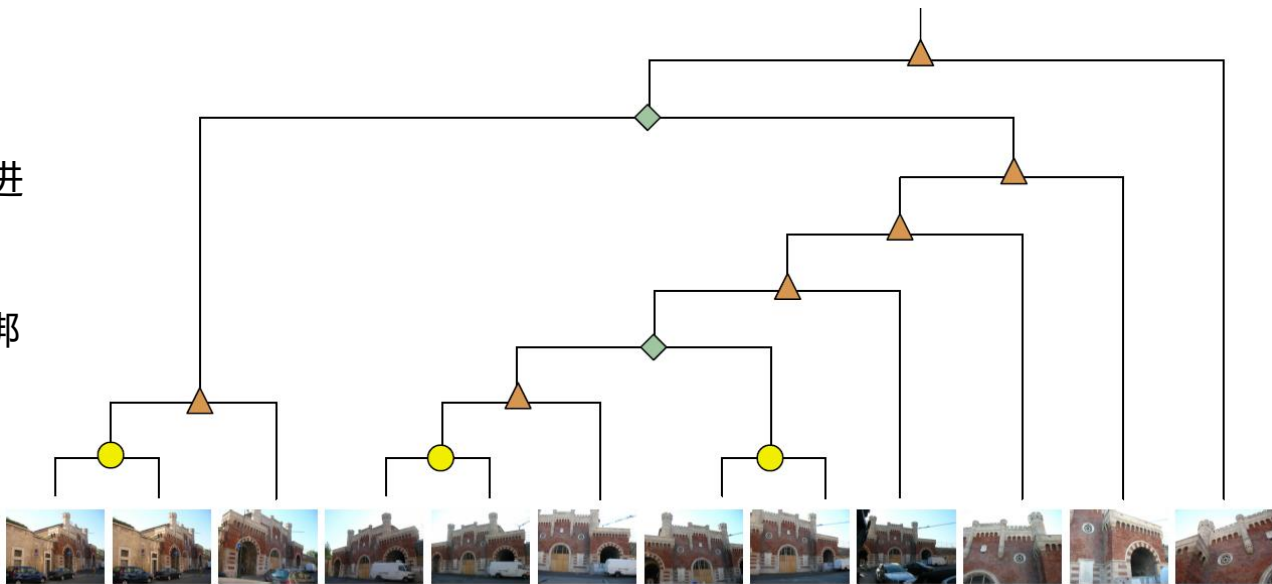
算法流程



分层式运动恢复结构

创建树状图

- 采用凝聚聚类的方式
- 根据图像内容的重合程度进行聚类
- 树状图用来引导分层的捆绑调整

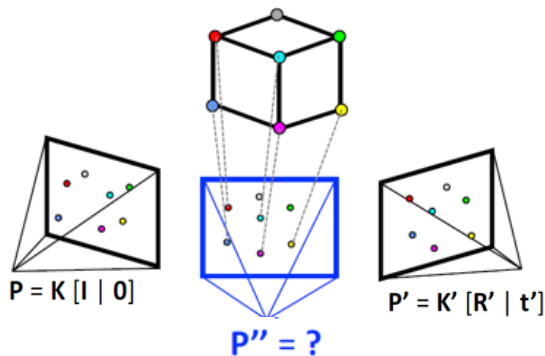


分层式运动恢复结构

三种基本操作

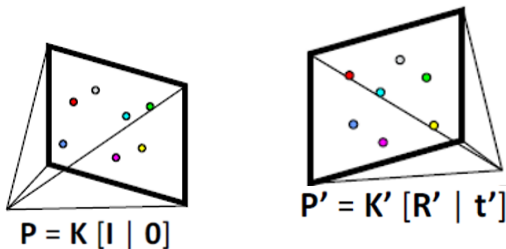
添加一个新的视角

- ✓ PnP恢复相机姿态
- ✓ 三角量测重建三维点



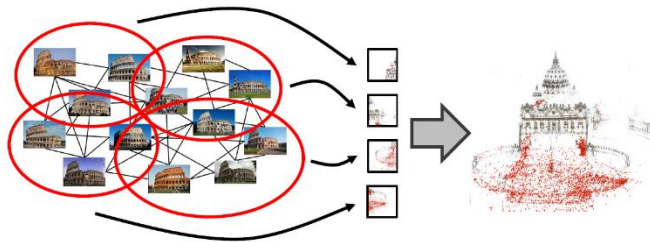
创建一个局部模型

- ✓ 计算相对姿态
- ✓ 三角量测重建三维点



融合两个局部模型

- ✓ 3D-3D 刚体变换

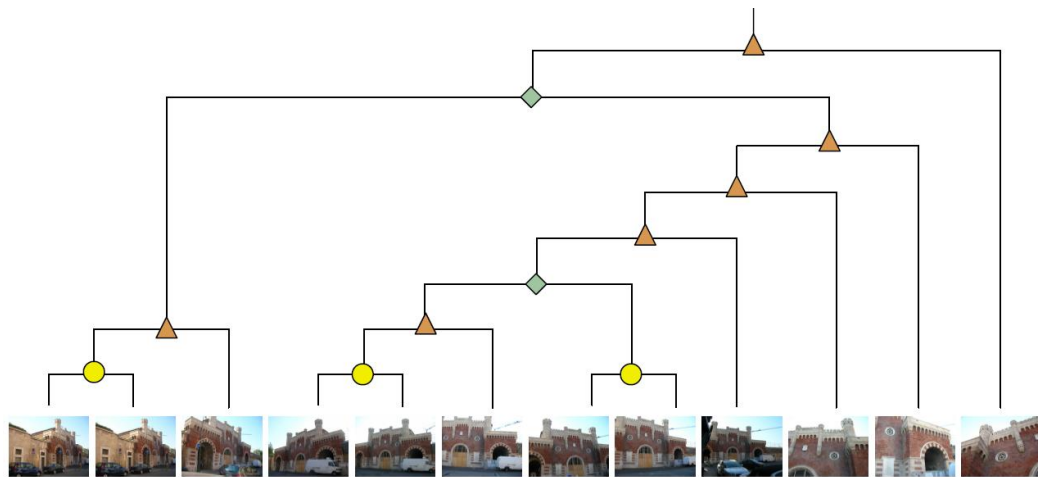


分层式运动恢复结构

● 创建一个新的局部模型 + BA

▲ 添加一个新的视角 + BA

◆ 融合两个局部模型 + BA



计算效率高,

算法稳定, 不依赖初始相机的选取

避免了大场景中的误差积累和漂移

规则的设计比较复杂

三种运动恢复结构的比较

方法	效率	鲁棒性	准确率
增量式	—	++	+
全局式	+	+	++
分层式	++	—	—

✓ 几种BA的形式

- ✓ 同时优化相机和三维点
- ✓ 优化相机
- ✓ 只优化三维点
- ✓ 单目相机

✓ 运动恢复结构

- ✓ 增量运动恢复结构 (Incremental SFM)
- ✓ 全局运动恢复结构 (Global SFM)
- ✓ 分层式运动恢复结构 (Hierarchical SFM)

✓ 运动恢复结构的几个问题

- 自标定的方法获取初始值

如 VisualSFM

- 从可交换图像文件中读取EXIF初始值

如 Bundler

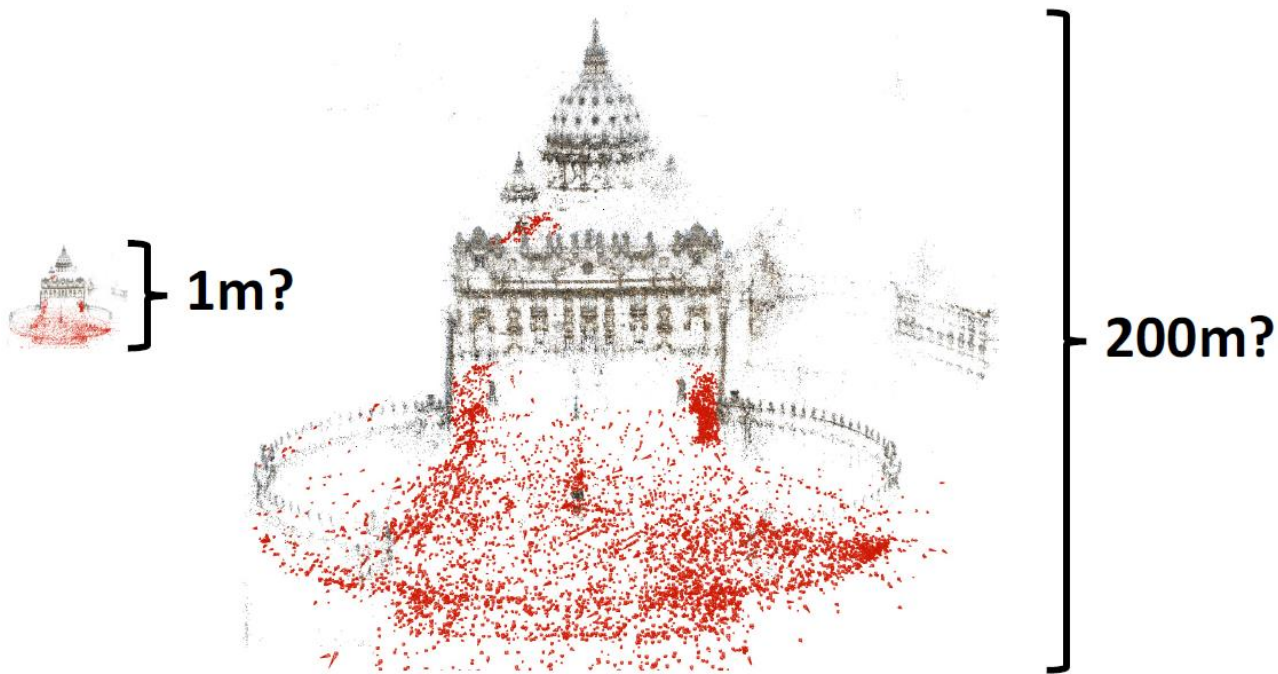
Exif



Exif是一种图像文件格式，它的数据存储与JPEG格式是完全相同的。实际上Exif格式就是在JPEG格式头部插入了数码照片的信息，包括拍摄时的光圈、快门、白平衡、ISO、焦距、日期时间等各种和拍摄条件以及相机品牌、型号、色彩编码、拍摄时录制的声音以及GPS全球定位系统数据、缩略图等。你可以利用任何可以查看JPEG文件的看图软件浏览Exif格式的照片，但并不是所有的图形程序都能处理Exif信息。

尺度不确定性

需要GPS进行地理标注



标准的SFM只能重建场景中的静态物体



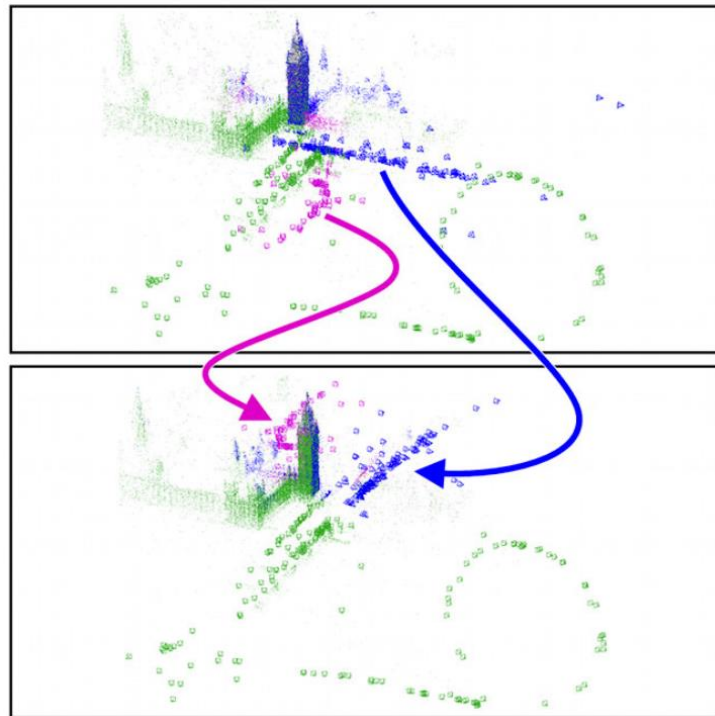
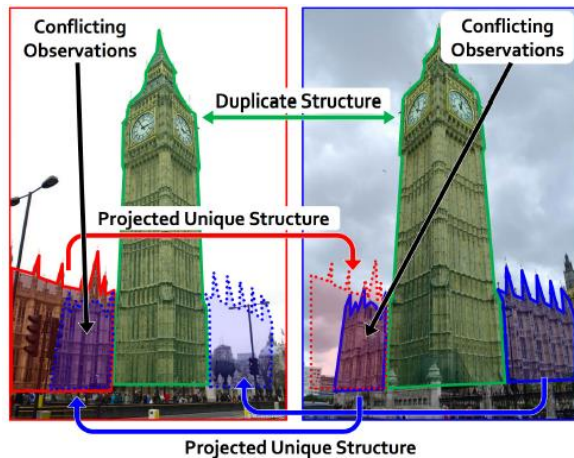
重建动态物体的一些工作



Zheng et al. 2015, "Sparse Dynamic 3D Reconstruction from Unsynchronized Videos"

Ji et al. 2014, "3D Reconstruction of Dynamic Textures in Crowdsourced Data"

重复结构



非朗伯面

找不到足够的匹配特征点



1. 阅读两篇增量SFM相关的参考文献

<https://www.gcc.tu-darmstadt.de/media/gcc/papers/Fuhrmann-2014-MVE.pdf>

<https://courses.cs.washington.edu/courses/cse576/10sp/notes/phototourism.pdf>

2. Coding 参见Task3.pdf



深蓝学院
shenlanxueyuan.com

感谢聆听！

Thanks for Listening