SCHOOL OF ENGINEERING SCIENCE

華中科技大學

HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

A Brief Introduction to

Structure from Motion

Zhang Yedi

2020.8.19 Wed.

CONTENTS

华中科技大学
HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

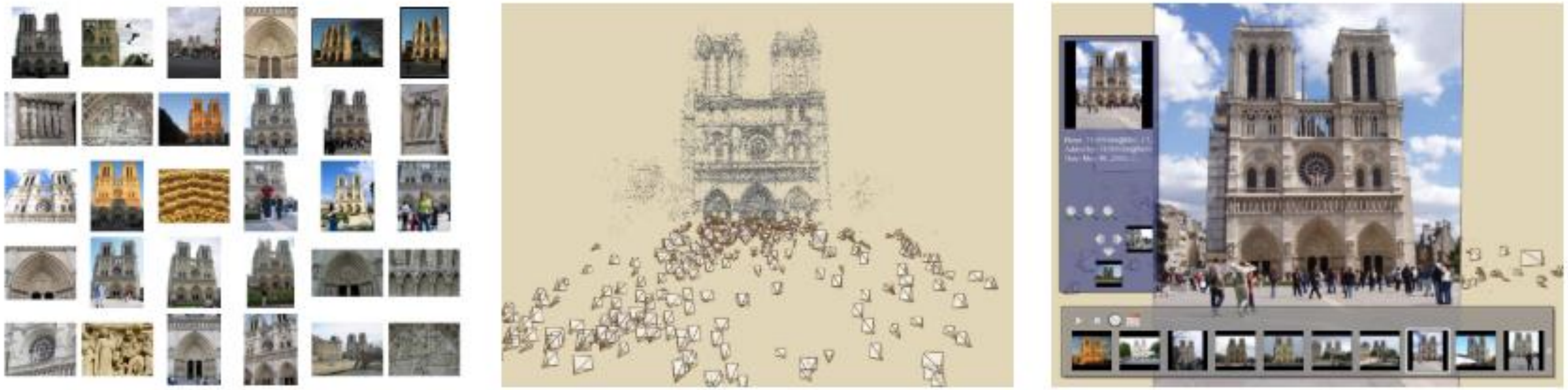
Origin：Snavely, Seitz, & Szeliski, 2006

Photo Tourism: Exploring Photo Collections in 3D

Noah Snavely
University of Washington

Steven M. Seitz
University of Washington

Richard Szeliski
Microsoft Research

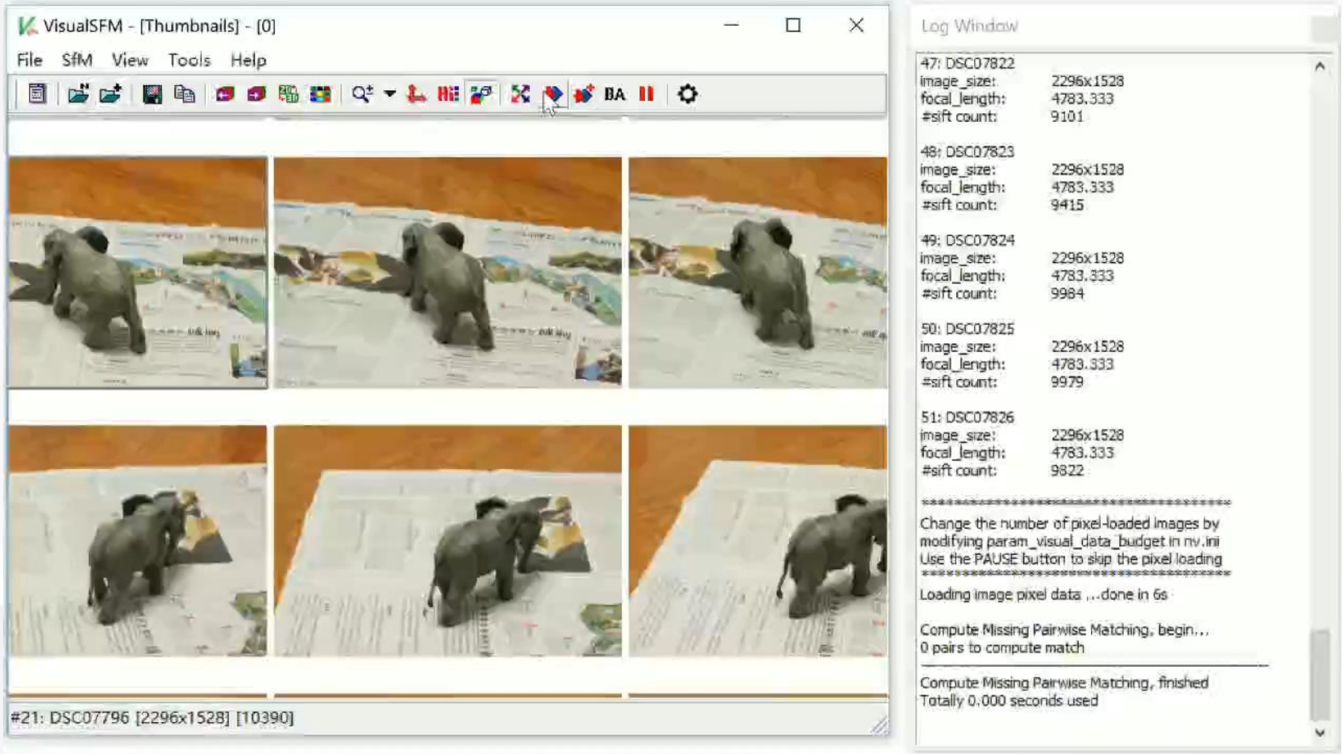


(a) (b) (c)

Figure 1: Our system takes unstructured collections of photographs such as those from online image searches (a) and reconstructs 3D points and viewpoints (b) to enable novel ways of browsing the photos (c).

Definition

- SfM = 利用运动的相机拍摄的图像生成稀疏3D点云



Reproduced results via [Changchang Wu, VisualSFM]

基本概念

- SfM = 利用运动的相机拍摄的图像生成3D点云
- 输入：

→ 图像及匹配的一系列特征点的像素坐标 $p_{ij} = (x_{ij}, y_{ij})$

- 输出：

→ Structure = 场景的3D点云

每个特征点 p_i的3D坐标X_i

→ Motion = 相机位置和朝向

相机外参t_j, R_j, (相机内参K_j)

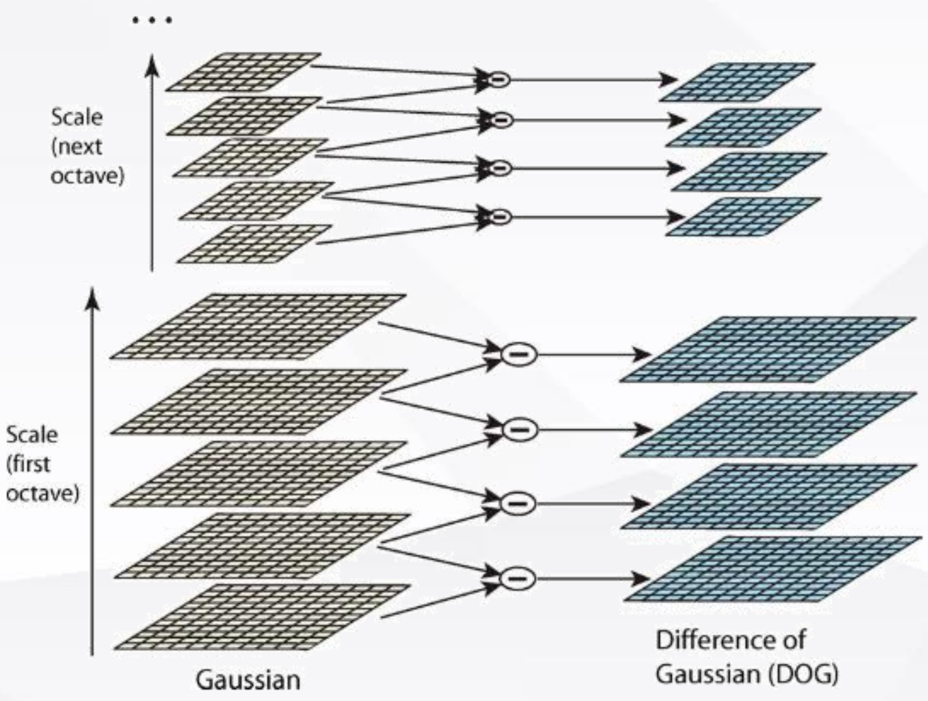
具体步骤

- 1 特征点检测与匹配
- 2 外极几何图构造
- 3 相机位姿和场景结构估计

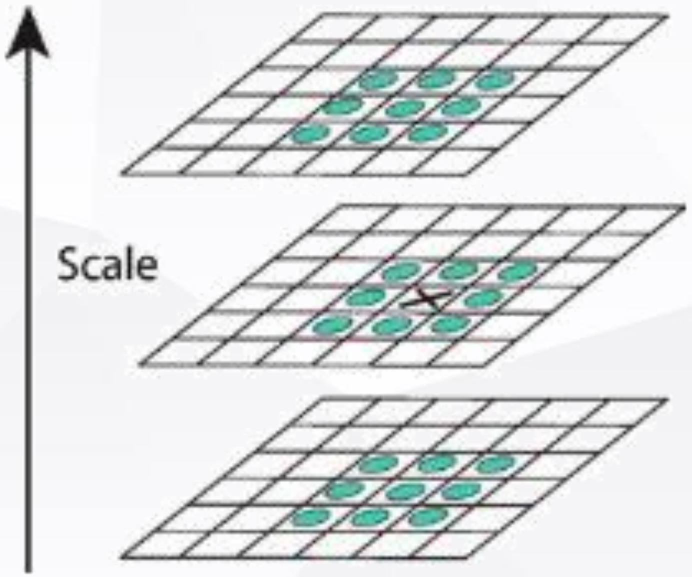
步骤1：特征点检测与匹配

- 特征点检测：SIFT, SURF, FAST, BRIEF, ORB, etc

①建立高斯差分金字塔



②确定关键点位置



步骤1：特征点检测与匹配

- 特征点检测：SIFT, SUFT, FAST, ORB, etc

③确定关键点方向

④创建关键点描述符

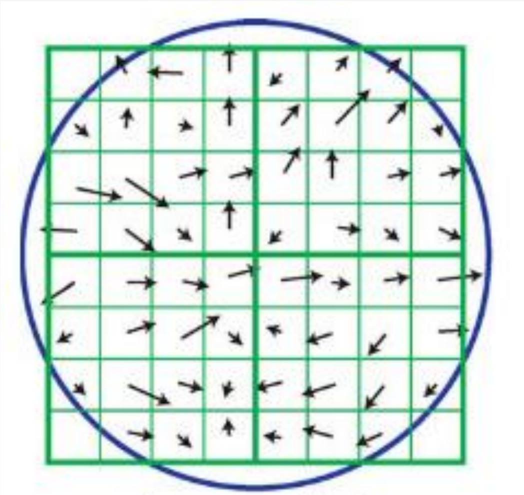
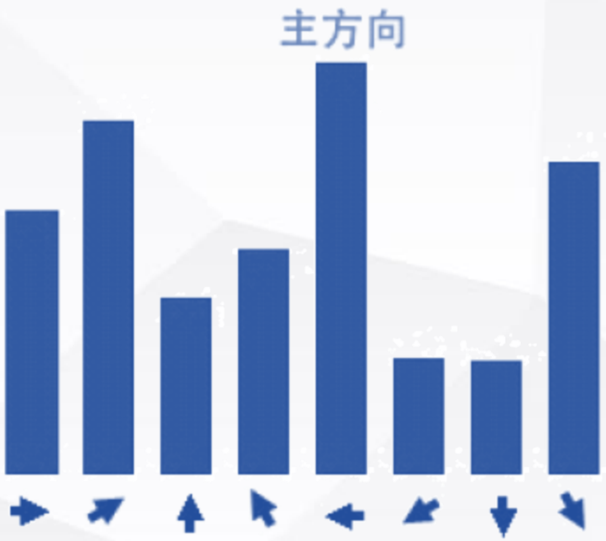
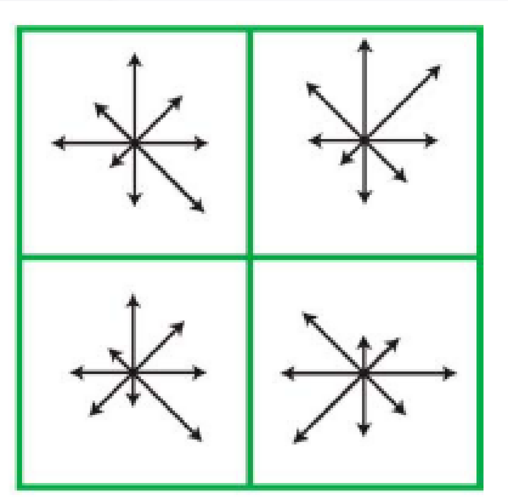


Image gradients





Keypoint descriptor

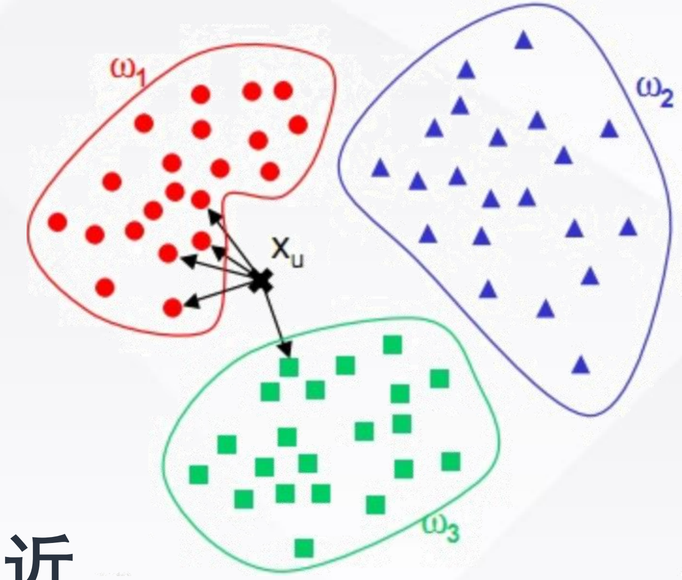
步骤1：特征点检测与匹配

- 特征点匹配：2张图片间的匹配

➢以KD树为数据结构，计算最近邻

➢比值约束：最近邻/次近邻小于一定阈值

➢外极几何约束：匹配特征点在对应的外极线附近

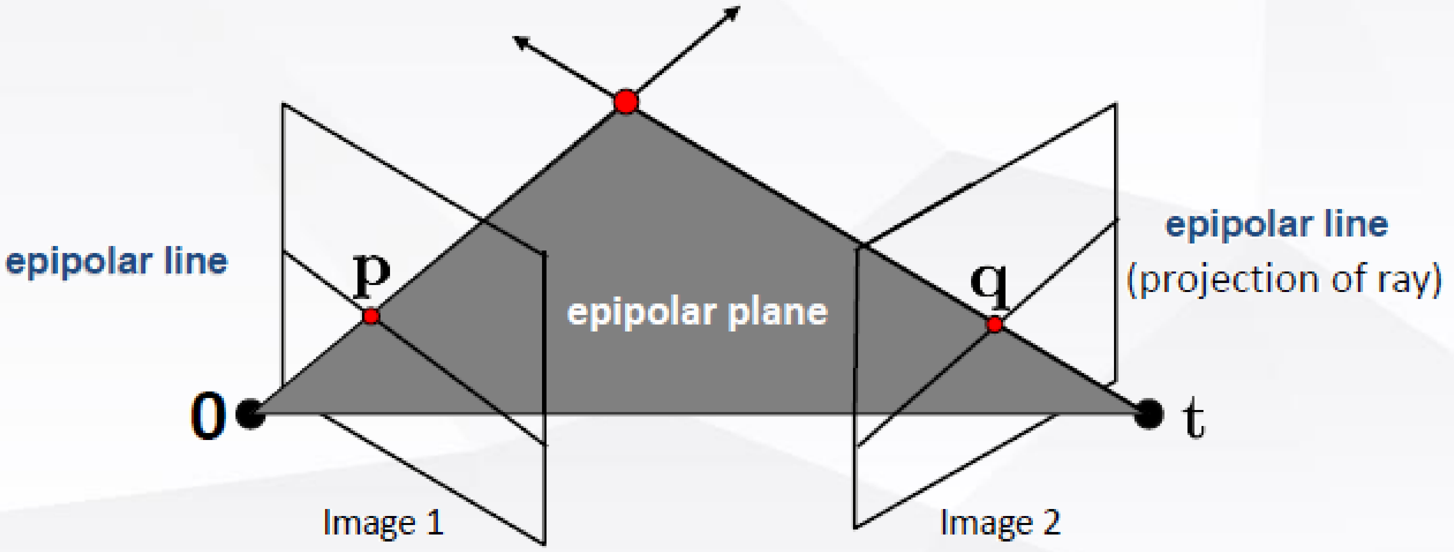


步骤2：外极几何图构造

- 目的：确定步骤3（摄像机位姿和场景结构估计）的初值
- 初值：通过重建2张图片，粗略给出3D点坐标$\boldsymbol{X}_i$和相机参数R_j, t_j, K_j

步骤2：外极几何图构造

- 初值：通过重建2张图片，粗略给出3D点坐标X_i和相机参数R_j, t_j, K_j
- 求解：8点法求本质矩阵$\boldsymbol{E}$，SVD分解出$\boldsymbol{R}, \boldsymbol{t}$，三角化求出3D点坐标$\boldsymbol{X}_i$



$$q^T \underbrace{K^{-T} \overbrace{t^\wedge R}^{\text{本质矩阵}\boldsymbol{E}} K^{-1}}_{\text{基础矩阵}\boldsymbol{F}} p = 0$$

性质：$q^T K^{-T} \boldsymbol{E} K^{-1} p = 0$

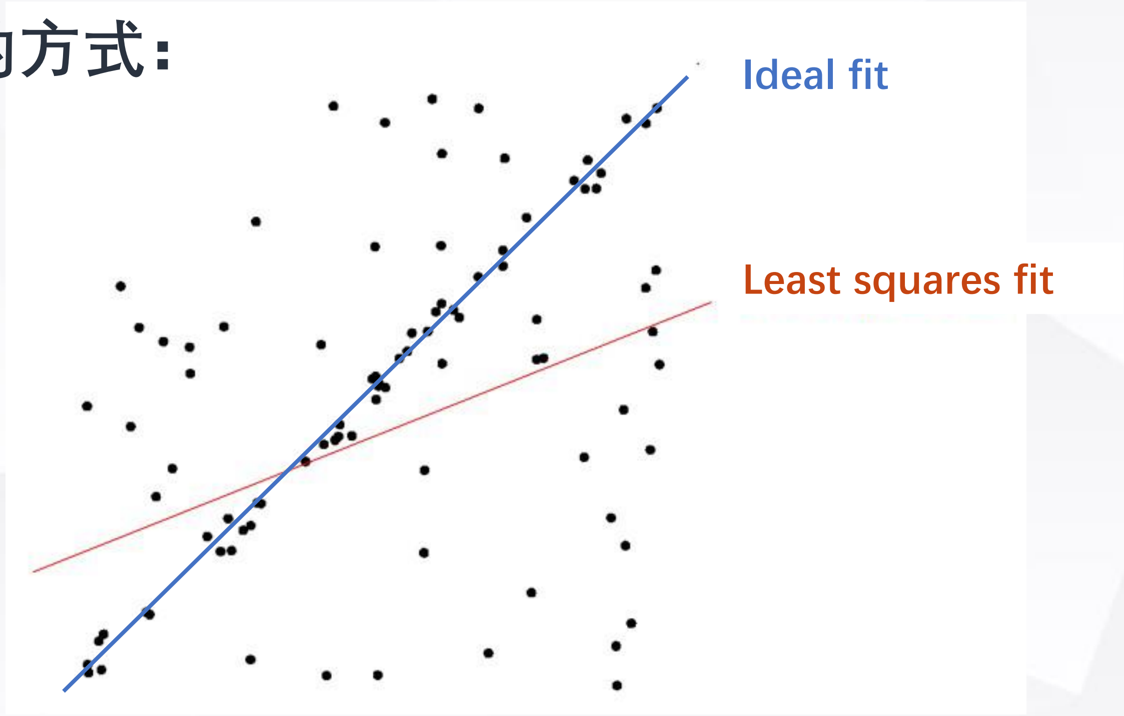
带入8对特征点求解本质矩阵

步骤2：外极几何图构造

- 初值：通过重建2张图片，粗略给出3D点坐标X_i和相机参数R_j, t_j, K_j
- 求解：8点法求本质矩阵$\boldsymbol{E}$，SVD分解出$\boldsymbol{R}, \boldsymbol{t}$，三角化求出3D点坐标$\boldsymbol{X_i}$

挑选8个点的方式：

RANSAC



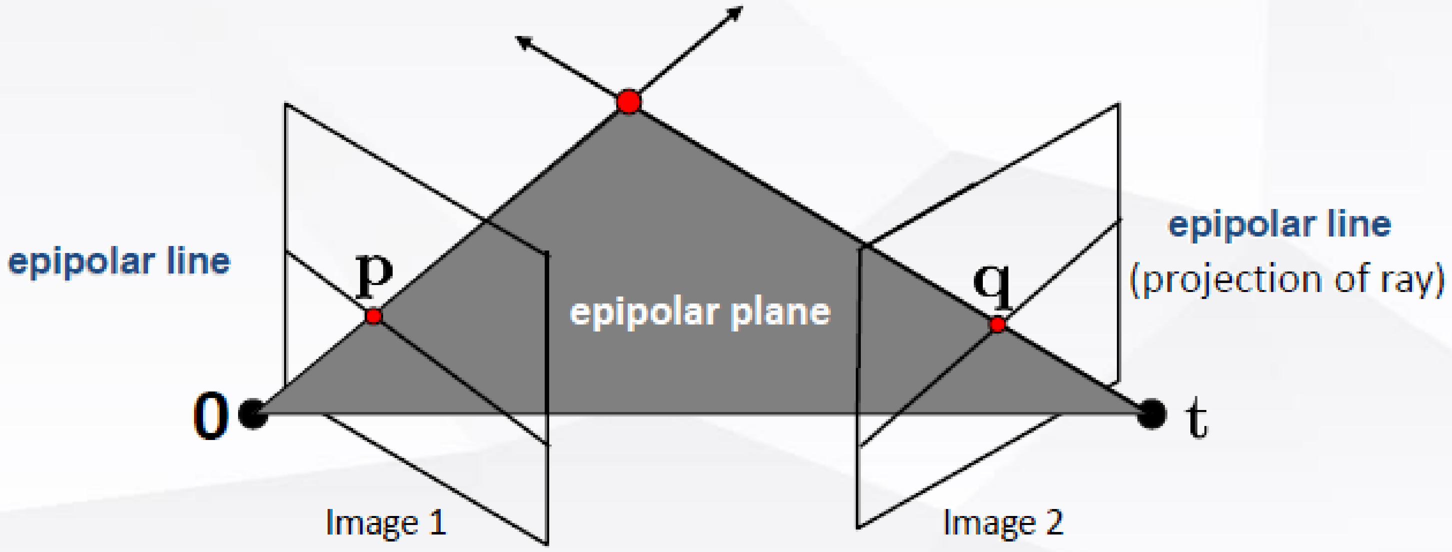
$$q^T \underbrace{K^{-T} \overbrace{\boxed{t^\wedge R}}^{\text{本质矩阵}\boldsymbol{E}} K^{-1}}_{\text{基础矩阵F}} p = 0$$

性质：$q^T K^{-T} \boldsymbol{E} K^{-1} p = 0$

带入8对特征点求解本质矩阵

步骤2：外极几何图构造

- 初值：通过重建2张图片，粗略给出3D点坐标X_i和相机参数R_j, t_j, K_j
- 求解：8点法求本质矩阵E，SVD分解出R, t，三角化求出3D点坐标X_i



本质矩阵：$E = t^{\wedge} R$

SVD分解：$E = U\Sigma V^T$

其中奇异值矩阵：

$$\Sigma = diag(\sigma, \sigma, 0)$$

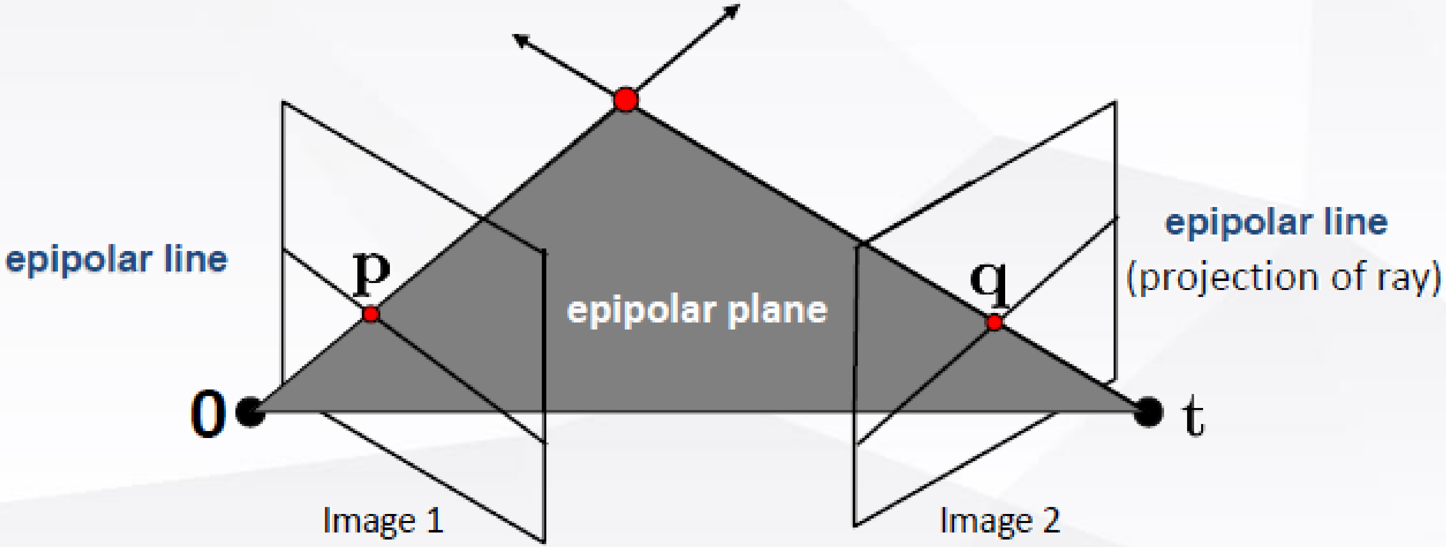
则得相机位姿：

$$t^{\wedge} = U r \Sigma U^T$$

$$R = U r \Sigma V^T$$

步骤2：外极几何图构造

- 初值：通过重建2张图片，粗略给出3D点坐标X_i和相机参数R_j, t_j, K_j
- 求解：8点法求本质矩阵E，SVD分解出R, t，三角化求出3D点坐标X_i



一对特征点的坐标满足：

$$d_1 p = d_2 R q + t$$

t, R已知。则由

$$d_1 p^\wedge p = 0 = d_2 p^\wedge R q + p^\wedge t$$

求得深度，从而获知3D点坐标

步骤3：摄像机位姿和场景结构估计

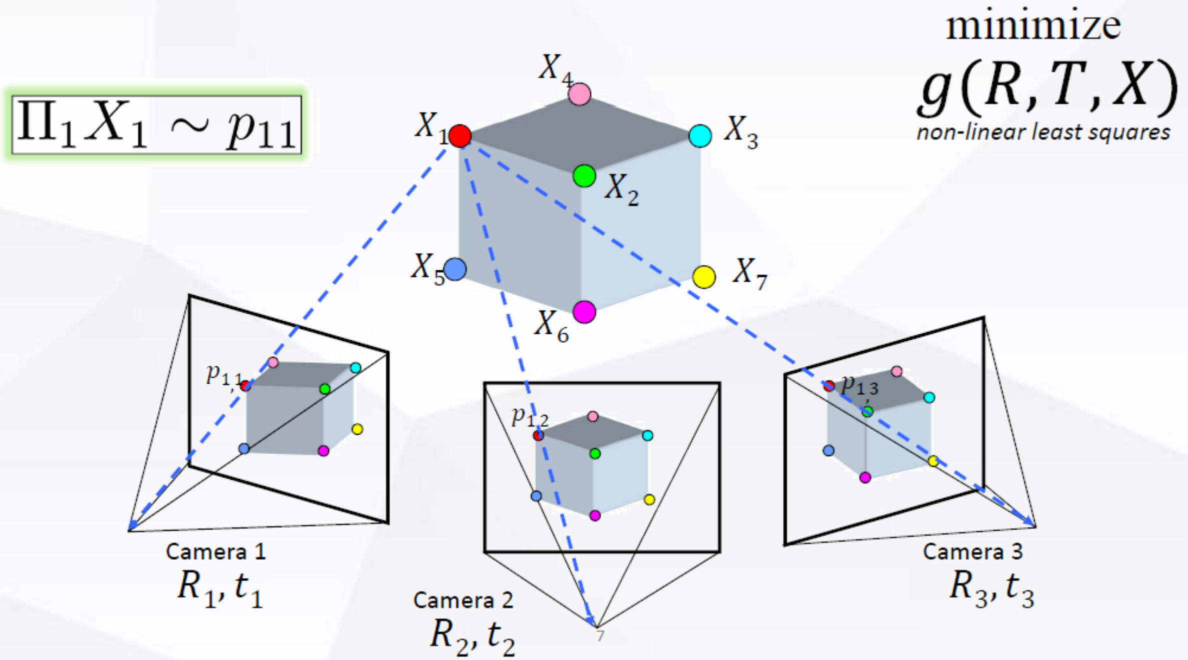
- 问题的描述：**最小化重投影误差**

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & u \\ 0 & f & v \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{K}\left[\mathbf{R} | \mathbf{t}\right] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

图像点　相机内参数　相机旋转　相机平移　空间点

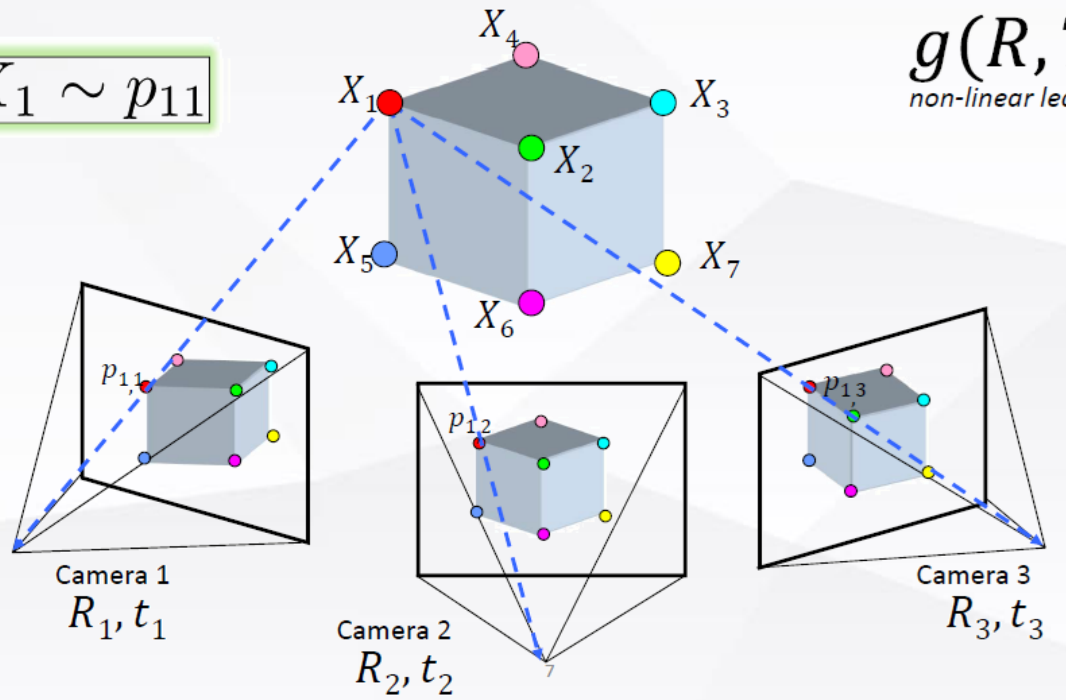
相机硬件　相机位姿　场景结构



步骤3：摄像机位姿和场景结构估计

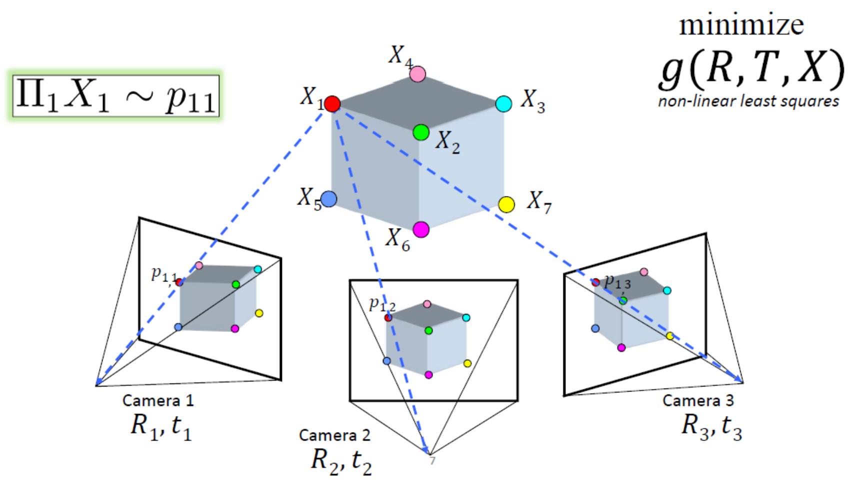
- 问题的描述：**最小化重投影误差**

$$\boxed{\Pi_1 X_1 \sim p_{11}}$$



$$\underset{\text{non-linear least squares}}{\text{minimize}}\; g(R,T,X) = \sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij} \cdot \left\| P(R_j, t_j, X_i) - \begin{bmatrix} x_{ij} \\ y_{ij} \end{bmatrix} \right\|^2$$

步骤3：摄像机位姿和场景结构估计



- 问题描述：最小化重投影误差

$$g(R,T,X)=\sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij}\cdot\left\|P\left(R_j,t_j,X_i\right)-\begin{bmatrix}x_{ij}\\ y_{ij}\end{bmatrix}\right\|^2$$

- 问题规模：$[\boldsymbol{n}\times(3+3+3+2)+\boldsymbol{m}\times 3]$ 个自变量
- 高维非线性最小二乘问题，做非线性优化
- 求解工具：Levenberg-Marquardt迭代法

步骤3：摄像机位姿和场景结构估计

- 求解工具：Levenberg-Marquardt迭代法
- 首先介绍Gauss-Newton迭代法

$$g(R,T,X)=\sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij}\cdot\left\|P\left(R_j,t_j,X_i\right)-\begin{bmatrix}x_{ij}\\ y_{ij}\end{bmatrix}\right\|^2=\sum_{i,j} e_{ij}^2(R,T,X)$$

$$g(R,T,X)=\sum_{i,j}[e_{ij}(P_0)+J_{ij}\Delta]^2\approx c+2b^T\Delta+\Delta^T H\Delta$$

- 对Δ求偏导并置为0，即：$\boldsymbol{H\Delta}+\boldsymbol{b}=\mathbf{0}$
- 赋值$\boldsymbol{P}\leftarrow\boldsymbol{P}+\Delta$，进行下一轮迭代

步骤3：摄像机位姿和场景结构估计

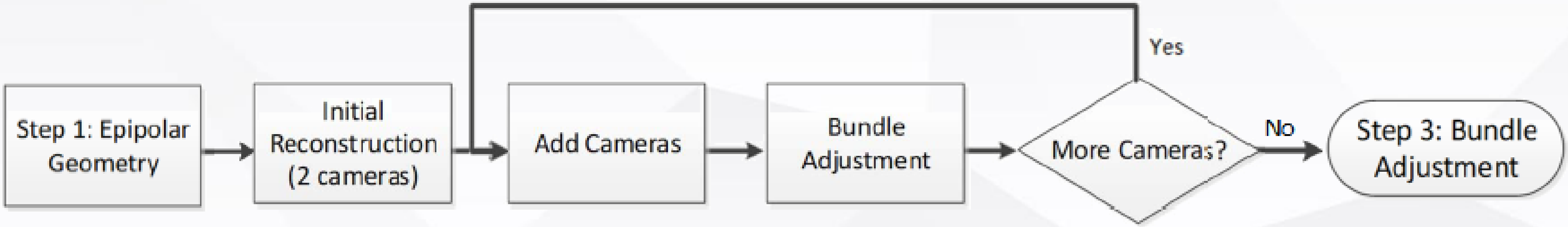
- 求解工具：Levenberg-Marquardt迭代法
- Gauss-Newton迭代法与梯度下降法的综合

$$g(R,T,X)=\sum_{i=1}^{m}\sum_{j=1}^{n}w_{ij}\cdot\left\|P\left(R_j,t_j,X_i\right)-\begin{bmatrix}x_{ij}\\y_{ij}\end{bmatrix}\right\|^2=\sum_{i,j}e_{ij}^2(R,T,X)\approx c+2b^T\Delta+\Delta^T H\Delta$$

- 对Δ求偏导，加入阻尼系数，再置为0，即：$(\boldsymbol{H}+\boldsymbol{\lambda I})\Delta+\boldsymbol{b}=\mathbf{0}$
- 若误差，接受Δ同时减小λ；若误差，不接受Δ同时增大λ
- 赋值$\boldsymbol{P}\leftarrow\boldsymbol{P}+\Delta$，进行下一轮迭代

步骤3：摄像机位姿和场景结构估计

- 光束平差法(Bundle Adjustment)：增量式，全局式，混合式



与开源API

- SIFT (SURF, FAST, BRIEF, ORB …)
- RANSAC
- Bundle adjustment (SBA, PBA, G2O, Ceres…)

与Simultaneous Localization and Mapping (SLAM)

- VSLAM：对具有时间先后顺序的图像序列实时求解SfM问题
- 侧重点不同，故优化的方向不同
- 数据源可能不同，单目、双目、RGB-D

三维重建的开源与商业软件

✓开源SfM软件

- OpenMVG 3030☆
- Colmap 2.3k☆
- OpenSfM (Python) 1.8k☆
- Bundler 1.2k☆
- MVE 669☆
- TheiaSfM 615☆
- MICMAC 292☆
- MAP-Tk

✓商业软件

- Pix4DMapper
- ContextCapture
- PhotoMesh
- Photoscan
- Streetfactory
- RealityCapture

内容小结

- **简介**
- **具体步骤**

  特征点检测与匹配

  外极几何图构造

  相机位姿和场景结构估计
- **相关内容**

  与SLAM的关联

  相关开源与商业软件

参考论文

- Snavely, Noah, Steven M. Seitz, and Richard Szeliski. "Photo Tourism: Exploring Photo Collections in 3d." *ACM Transactions on Graphics* (2006). http://phototour.cs.washington.edu/.
- Snavely, Noah, Steven M. Seitz, and Richard Szeliski. "Modeling the World from Internet Photo Collections." *International Journal of Computer Vision* 80, no. 2 (2008): 189-210. https://doi.org/10.1007/s11263-007-0107-3. https://dx.doi.org/10.1007/s11263-007-0107-3.
- Lowe, David G. "Distinctive Image Features from Scale-Invariant Keypoints." *International Journal of Computer Vision* 60, no. 2 (2004): 91-110. https://doi.org/10.1023/b:visi.0000029664.99615.94. https://dx.doi.org/10.1023/B:VISI.0000029664.99615.94.
- Fischler, Martin A., and Robert C. Bolles. "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography." *Communications of the ACM* 24, no. 6 (1981): 381-95. https://doi.org/10.1145/358669.358692.
- Triggs, Bill, Philip F. McLauchlan, Richard I. Hartley, and Andrew W. Fitzgibbon. "Bundle Adjustment — a Modern Synthesis." 298-372: Springer Berlin Heidelberg, 2000.
- Maier-Hein, L., P. Mountney, A. Bartoli, H. Elhawary, D. Elson, A. Groch, A. Kolb, *et al.* "Optical Techniques for 3d Surface Reconstruction in Computer-Assisted Laparoscopic Surgery." Review. *Med Image Anal* 17, no. 8 (Dec 2013): 974-96. https://doi.org/10.1016/j.media.2013.04.003. https://www.ncbi.nlm.nih.gov/pubmed/23837969.

其他参考资料

- 《视觉SLAM十四讲：从理论到实践》，高翔、张涛等著
- 《基于图像点特征的多视图三维重建》，康来著
- CCCV2017讲习班《基于图像的大规模场景三维重建》，中科院申抒含、崔海楠
- 浙江大学谭平老师课程《计算机视觉》
- B站up主会飞的吴克

SCHOOL OF ENGINEERING SCIENCE

華中科技大學

HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Thank you!

Zhang Yedi

2020.8.19 Wed.