

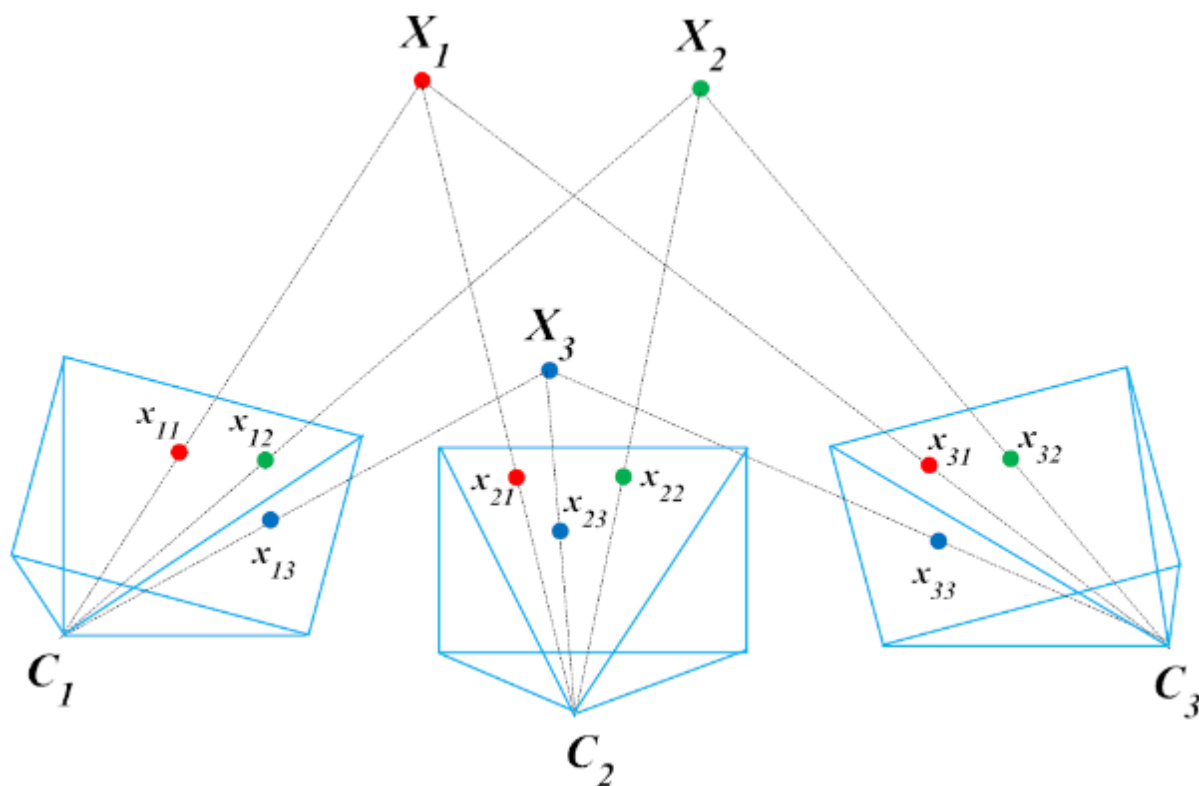
运动恢复结构

章国锋

浙江大学**CAD&CG**国家重点实验室



多视图几何



$$\mathbf{x}_{ij} = \pi(\mathbf{P}_i X_j)$$

投影函数 $\pi(x, y, z) = (x/z, y/z)$ $\mathbf{P}_i = \mathbf{K}_i[\mathbf{R}_i | \mathbf{T}_i]$

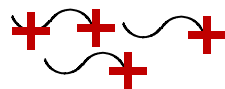
运动恢复结构

- 流程

- 特征跟踪

- 获得一堆特征点轨迹

$$\mathcal{X} = \{\mathbf{x}_i | i=1, \dots, m\}$$



- 运动恢复结构

- 求解相机参数和特征点轨迹的三维位置

$$\mathbf{x}_{ij} = \pi(\mathbf{P}_i X_j) \quad \mathbf{P}_i = \mathbf{K}_i[\mathbf{R}_i | \mathbf{T}_i]$$

$$E(\mathbf{P}_1, \dots, \mathbf{P}_m, X_1, \dots, X_n) = \sum_{i=1}^m \sum_j^n w_{ij} ||\pi(\mathbf{P}_i X_j) - \mathbf{x}_{ij}||^2$$

图像特征

- 图像中显著、容易区分和匹配的内容
 - 点
 - 角点
 - 线: 直线, 曲线,...
 - 边: 二维边, 三维边
 - 形状: 长方形, 圆, 椭圆, 球,...
 - 纹理
- 不变性
 - 视角不变(尺度, 方向, 平移)
 - 光照不变
 - 物体变形
 - 部分遮挡

Harris 角点检测

- 核心思想：统计图像梯度的分布
 - 平坦区域：梯度不明显
 - 边缘区域：梯度明显，方向一致
 - 角点区域：梯度明显，方向不一致

- 方法：

- 计算像素邻域的梯度二阶矩

$$H = \sum_{(u,v)} w(u,v) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

- 计算上述矩阵的角点响应指标

$$R = \det(H) - \alpha \cdot \text{trace}(H)^2$$

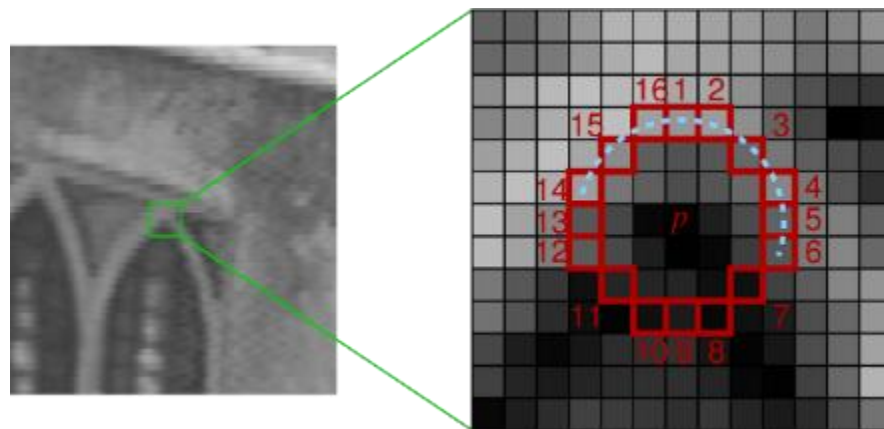
- 对 R 进行阈值过滤和非极大值抑制

FAST

- 核心思想
 - 角点与周围邻域内足够多的像素的灰度差异较大
- 通过少量像素点的比较，加速角点提取
- 考虑中心点周围的16个像素，设中心点亮度为 p
 - 如果有连续 n 个像素亮度都大于 $p+t$ ，或者都小于 $p-t$ (如图中的 14~16, 1 ~ 6)
 - 首先检查 1、5、9、13 四个位置，如果是角点，四个位置中应当有三个满足上面的条件

•

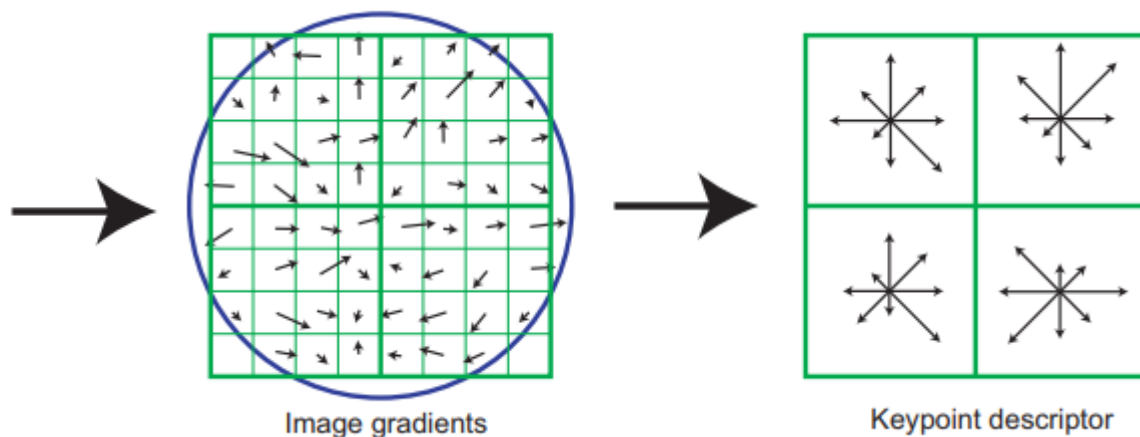
- 速度快，但对噪音不鲁棒



Edward Rosten, Tom Drummond. Machine Learning for High-Speed Corner Detection. ECCV (1) 2006: 430-443.

SIFT

- Scale-Invariant Feature Transform
- SIFT通过在不同级别的图像DoG上寻找极值点来确定特征点的位置和对应的尺度，后续的特征提取在与其尺度最邻近的图像DoG上进行。这使它有良好的尺度不变性。



David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60(2): 91-110 (2004).

More Invariant Features

- SIFT之后陆续出现了各种尺度不变特征描述量提取算法
 - 如 RIFT、GLOH、SURF等
 - 保证了一定的视觉不变性，又能很好地对抗噪声
- SURF
 - 使用了Haar小波卷积替代SIFT中的高斯核
 - 用积分图像进行了加速，使得计算速度达到SIFT的3~7倍
- ORB
 - 使用FAST提取特征点
 - 使用轻量级的二进制描述子
 - 由于其极快的提取速度得到了广泛使用。

特征提取



SIFT

极佳的尺度不变性，能一定程度上适应视角变化和亮度变化

SURF

能够处理严重的图像模糊，速度要高于SIFT，但精度不如SIFT

ORB

极快的提取速度，在实时应用中常用来替代SIFT

以上三种特征提取算法均在OpenCV中有实现

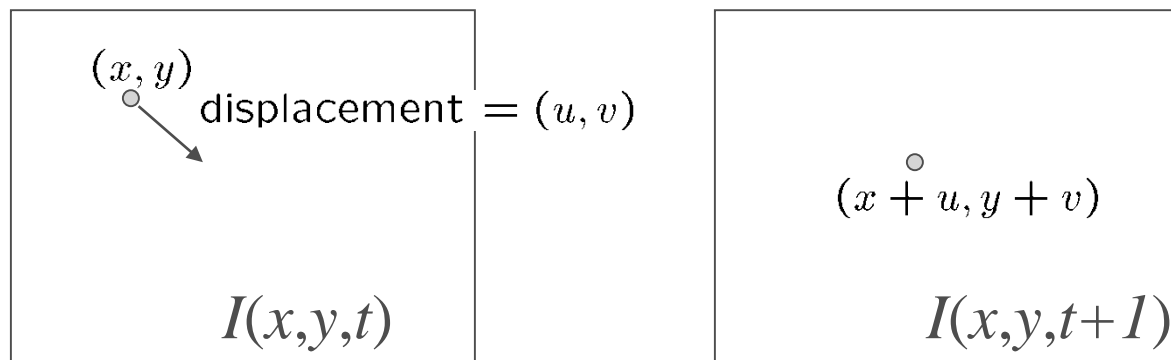
特征匹配

- 模板匹配
 - 直接在目标图像中寻找给定的图像块



特征匹配

在小运动假设下，可以采用 KLT 跟踪方法：




$$\begin{aligned} I(x,y,t) &= I(x+u, y+v, t+1) \\ &\approx I(x,y,t) + \boxed{I_x u + I_y v + I_t} \rightarrow \nabla I \cdot \begin{pmatrix} u \\ v \end{pmatrix} + I_t = 0 \end{aligned}$$

一个等式，两个未知量

特征匹配

进一步假设：相邻像素运动一致

$$\nabla I \cdot \begin{pmatrix} u \\ v \end{pmatrix} + I_t = 0 \quad (\text{单个像素})$$

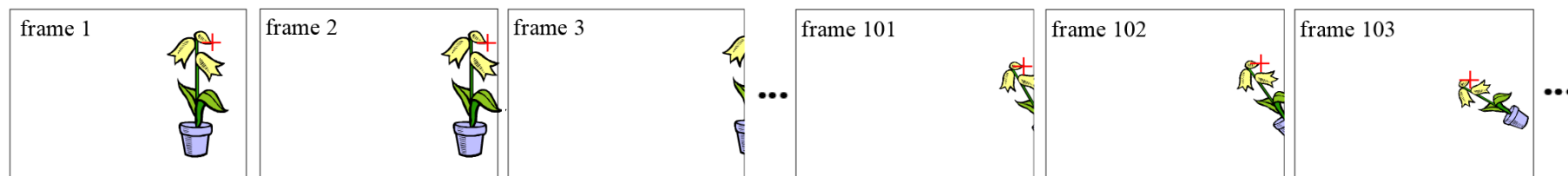

$$\begin{pmatrix} I_x(p_1) & I_y(p_1) \\ \vdots & \vdots \\ I_x(p_n) & I_y(p_n) \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} I_t(p_1) \\ \vdots \\ I_t(p_n) \end{pmatrix} = 0 \quad (\text{邻域窗口})$$

特征匹配

- 大运动情况下的匹配
 - 通过比较特征描述量的距离进行匹配
 - SIFT = 128 维、SURF = 64 维、ORB = 256bits
 - 暴力匹配
 - 快速最近邻匹配
 - OpenCV中提供了相应的匹配算法

Non-Consecutive Feature Tracking

- How to efficiently match the common features among different subsequences?



Non-Consecutive Feature Tracking



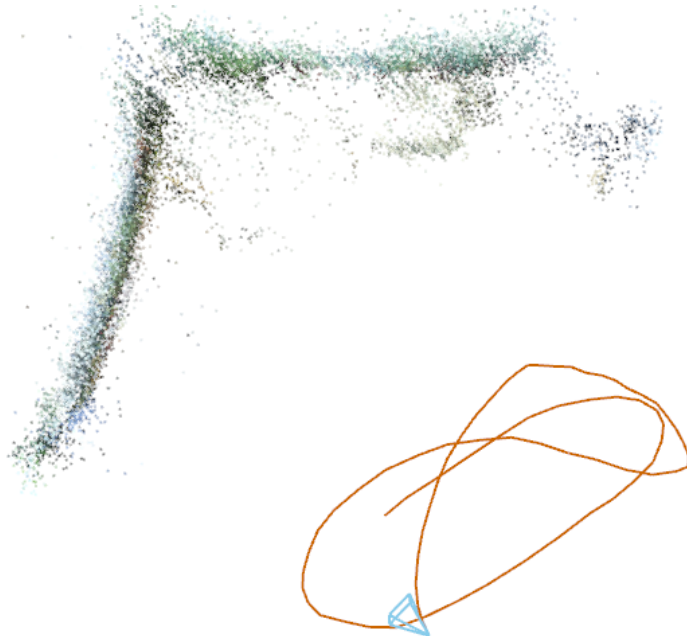
**Consecutive Feature
Tracking**



**Non-Consecutive
Track Matching**



**Structure from
Motion**



Framework Overview

1. Detect SIFT features over the entire sequence.
2. **Consecutive point tracking:**
 - 2.1 Match features between consecutive frames with descriptor comparison.
 - 2.2 Perform the second-pass matching to extend track lifetime.
3. **Non-consecutive track matching:**
 - 3.1 Use hierarchical k-means to cluster the constructed tracks.
 - 3.2 Estimate the matching matrix with the grouped tracks.
 - 3.3 Detect overlapping subsequences and join the matched tracks.

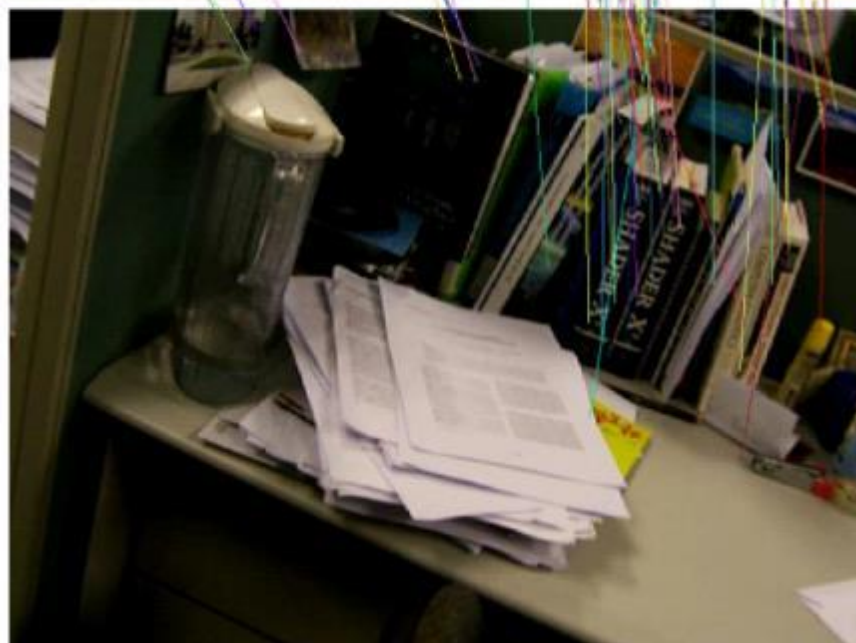
Two-Pass Matching

- SIFT Feature Extraction
- First-Pass Matching by Descriptor Comparison

$$c = \frac{\|\mathbf{p}(\mathcal{N}_1^{t+1}(\mathbf{x}_t)) - \mathbf{p}(\mathbf{x}_t)\|}{\|\mathbf{p}(\mathcal{N}_2^{t+1}(\mathbf{x}_t)) - \mathbf{p}(\mathbf{x}_t)\|}$$

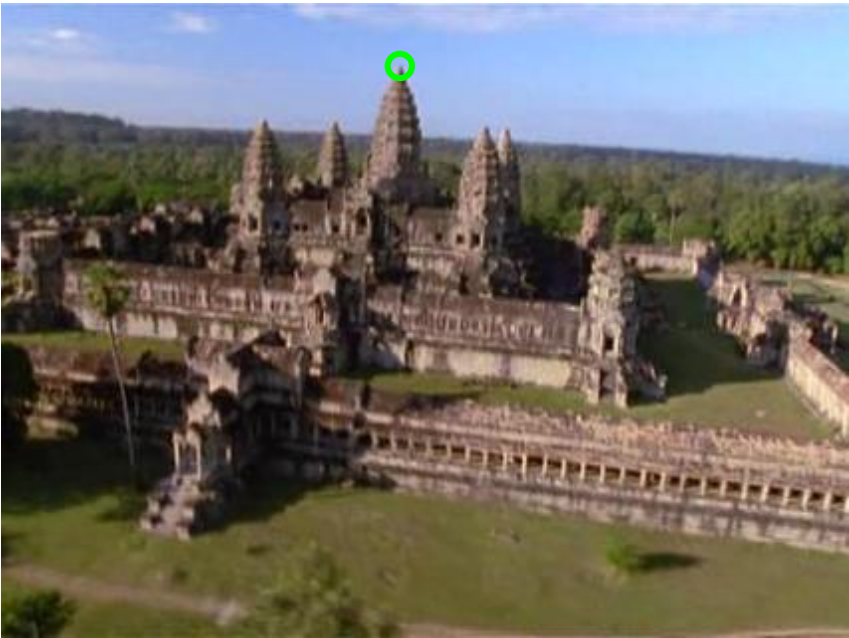
$c < \varepsilon$ Global distinctive





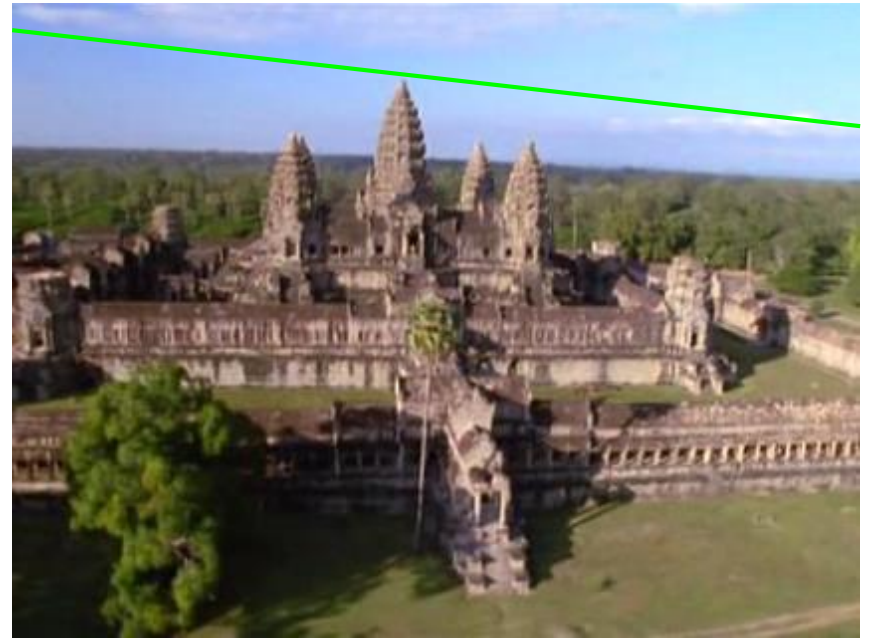
Two-View Geometry

3D???



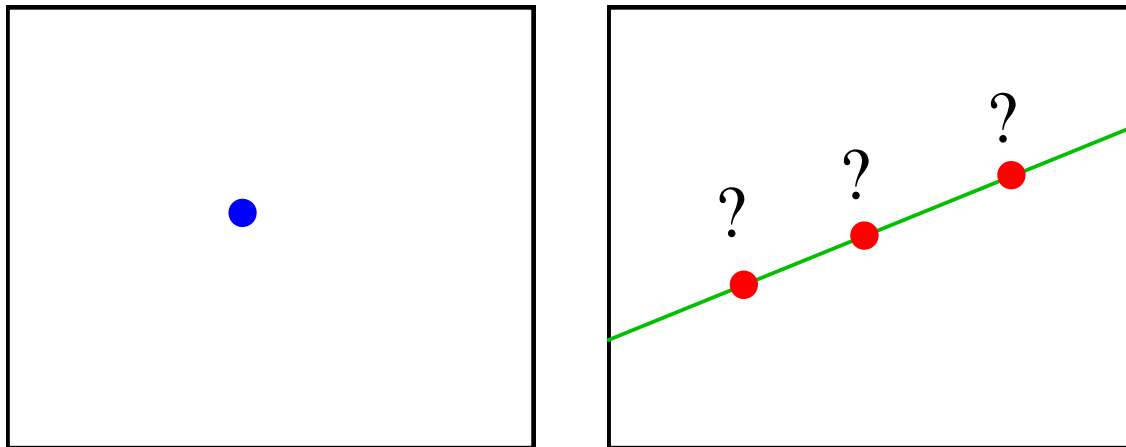
Two-View Geometry

3D: Epipolar Geometry



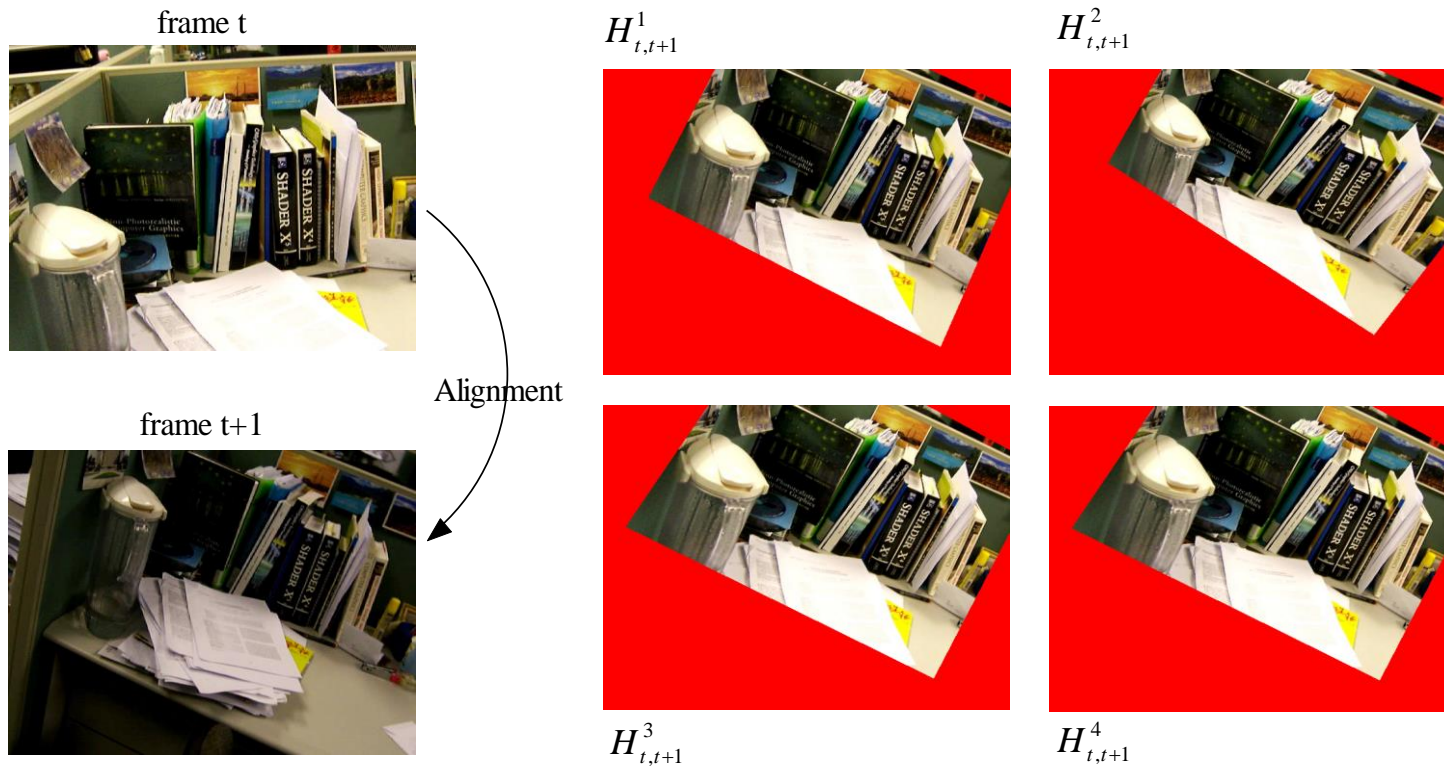
Not Enough!

- How to handle image distortion?
 - Naïve window-based matching becomes unreliable!
- How to give a good position initializaton?
 - Whole line searching is still time-consuming and ambiguous with many potential correspondences.



Second-Pass Matching by Planar Motion Segmentation

- Estimate a set of homographi $\{H_{t,t+1}^k | k = 1, \dots, N\}$
 - Using inlier matches in first-pass matching

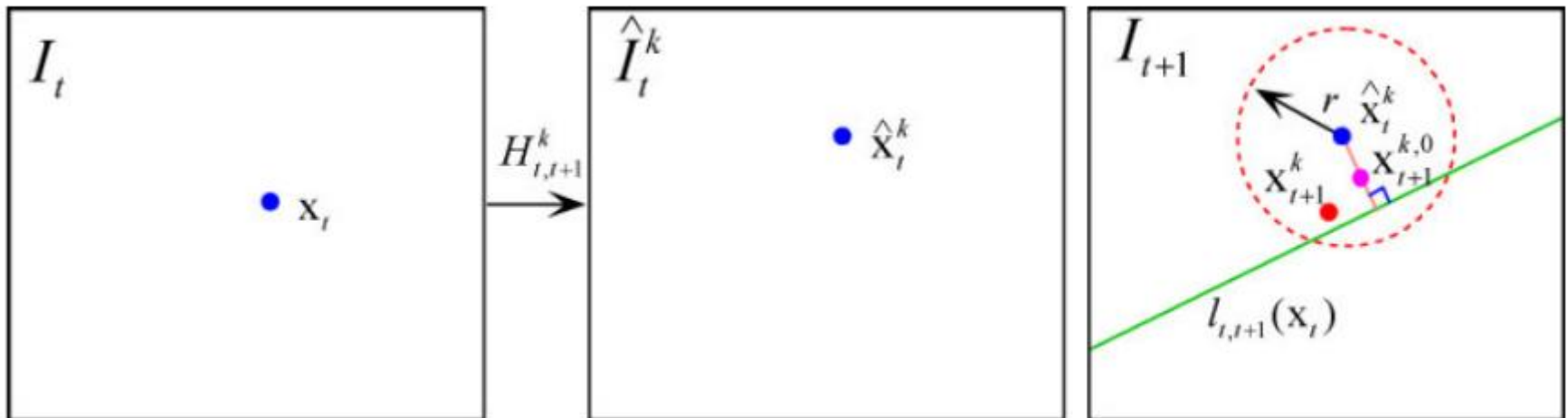


Second-Pass Matching by Planar Motion Segmentation

- Guided matching

$$S_{t,t+1}^k(\mathbf{x}_{t+1}^k) = \sum_{\mathbf{y} \in W} \|\hat{I}_t^k(\hat{\mathbf{x}}_t^k + \mathbf{y}) - I_{t+1}(\mathbf{x}_{t+1}^k + \mathbf{y})\|^2 +$$
$$\lambda_e d(\mathbf{x}_{t+1}^k, l_{t,t+1}(\mathbf{x}_t))^2 + \lambda_h \|\hat{\mathbf{x}}_t^k - \mathbf{x}_{t+1}^k\|^2$$

Epipolar constraint Homography constraint



Second-Pass Matching with Multi-Homographies



First-Pass Matching
(53 matches)



Direct Searching
(11 matches added)



Our Second-Pass Matching
(346 matches added)

Non-Consecutive track matching

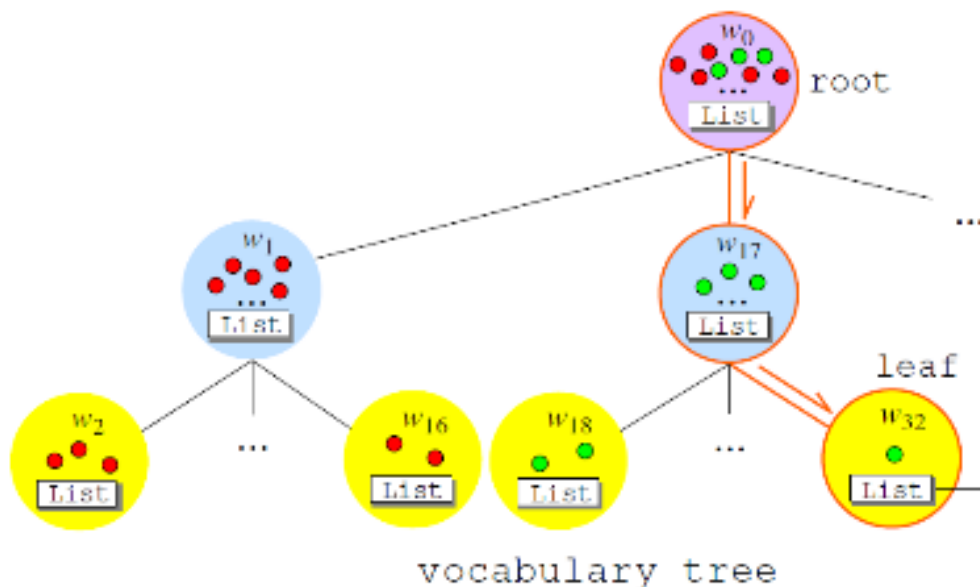
- Fast Matching Matrix Estimation
- Detect overlapping subsequences and join the matched tracks.

Fast Matching Matrix Estimation

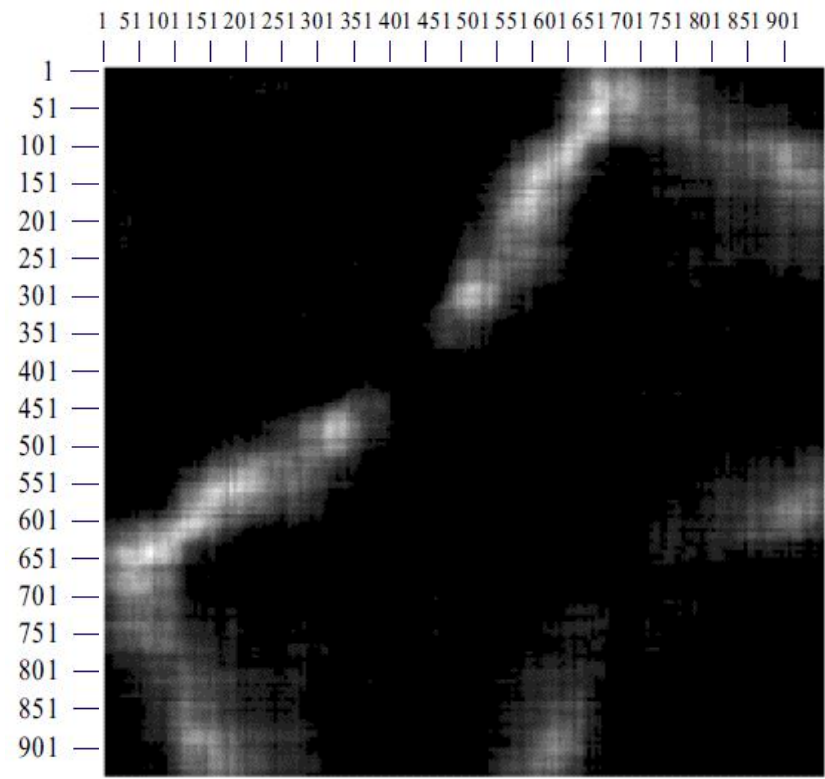
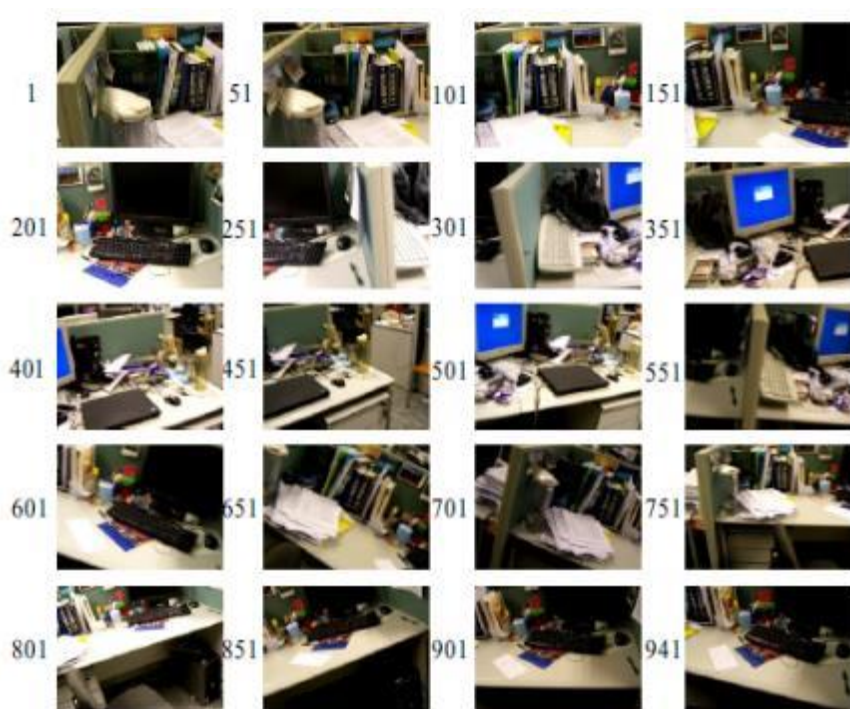
- Each track has a group of description vectors

$$\mathcal{P}_{\mathcal{X}} = \{\mathbf{p}(\mathbf{x}_t) | t \in f(\mathcal{X})\}$$

- Track descriptor $\mathbf{p}(\mathcal{X})$
- Use a hierarchical K-means approach to cluster the track descriptors

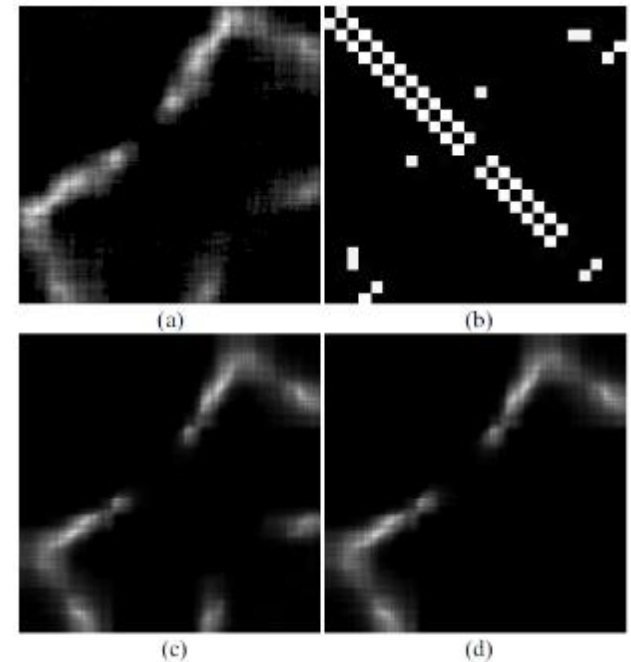
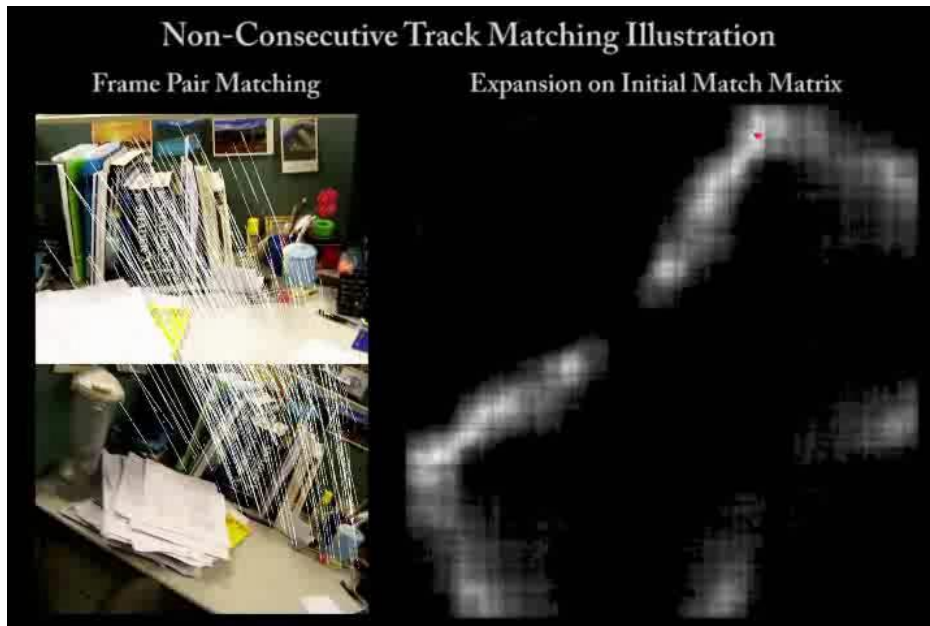


Fast Matching Matrix Estimation

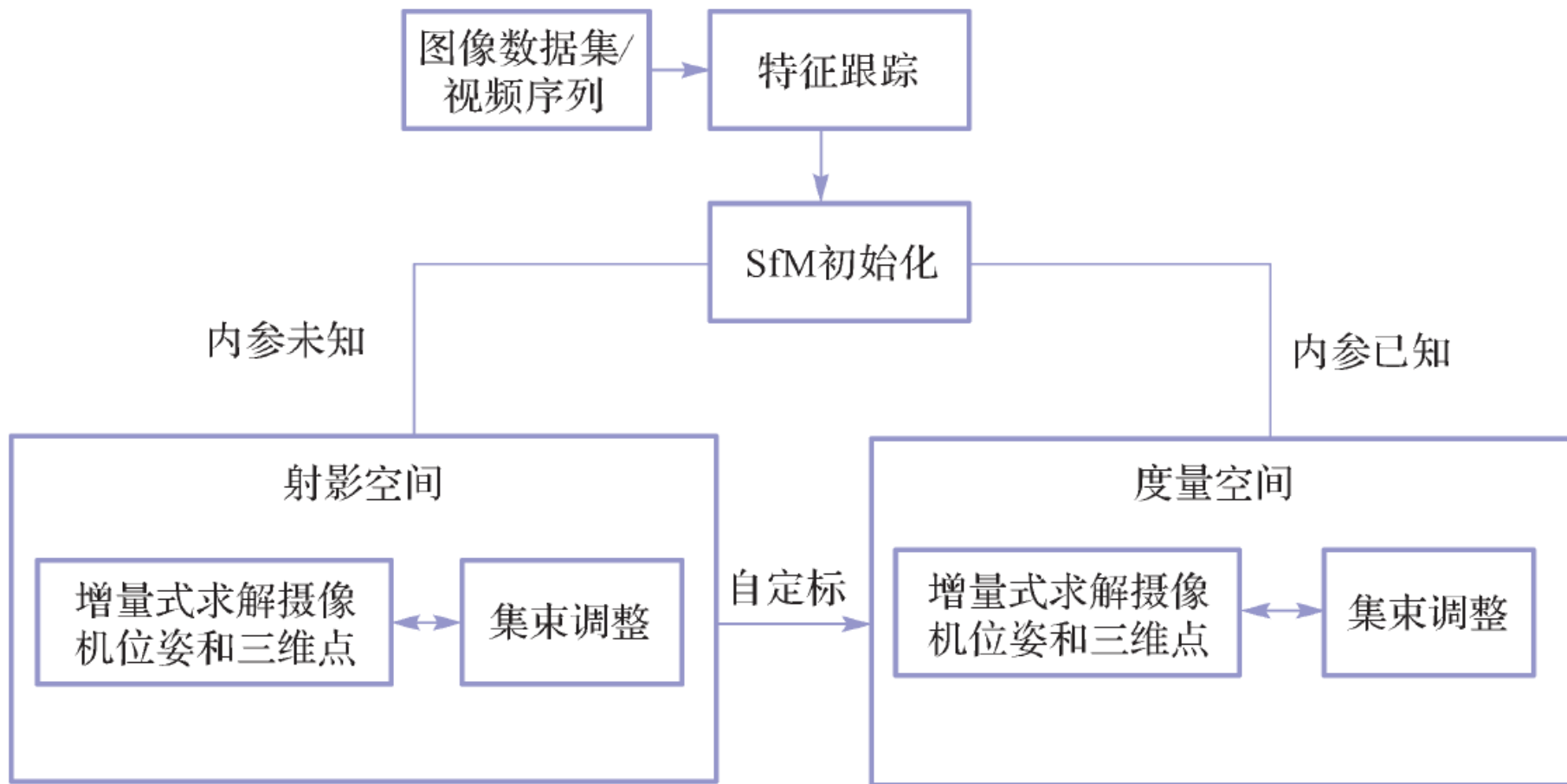


Non-Consecutive Track Matching

- Simultaneously Match Images and Refine Matching Matrix
 - Refine the matching matrix after matching the common features of the selected image pairs.
 - More reliably find the best matching images with the updated matching matrix.



常用的增量式SfM 系统框架



基于自定标的单序列增量式SfM 求解框架

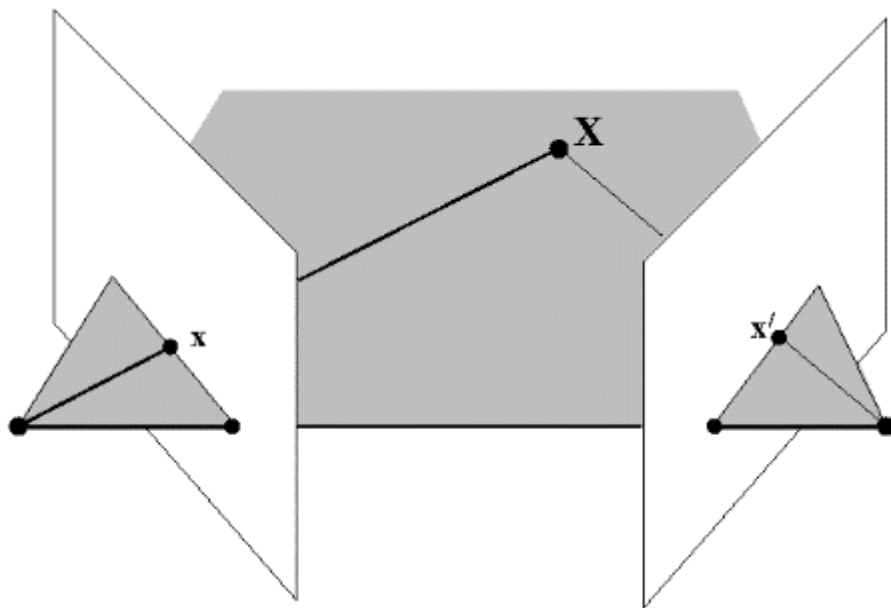
- 1.自动抽取特征点并匹配;
- 2.抽取关键帧组成关键帧序列;
- 3.初始化度量空间下的三维结构和运动:
 - 3.1 选择合适的三帧组进行射影重建的初始化;
 - 3.2 采用增量式求解, 并选择合适时机进行自定标, 将射影重建转换到度量重建;
- 4.对于每一个新加入求解的关键帧:
 - 4.1 初始化新求解帧的相机参数和相关的三维点;
 - 4.2 用局部集束调整算法对局部已经求解的结构和运动进行求精;
- 5.求解所有非关键帧的相机参数;
- 6.对整个序列恢复的结构和运动用集束调整进行最后优化。

三角化

- 已知 \mathbf{F} , 计算 \mathbf{P} 和 \mathbf{P}'

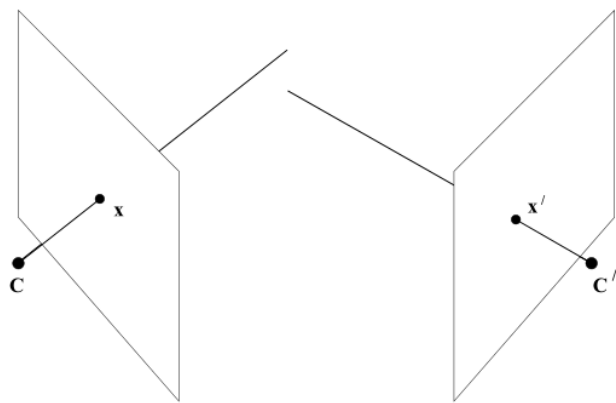
$$\mathbf{P} = [\mathbf{I} \mid \mathbf{0}] ; \mathbf{P}' = [[\mathbf{e}']_{\times} \mathbf{F} \mid \mathbf{e}'] = [\mathbf{M} \mid \mathbf{e}']$$

- 已知 \mathbf{x} 和 \mathbf{x}'
- 计算 \mathbf{X} : $\mathbf{x} = \mathbf{P}\mathbf{X}$ $\mathbf{x}' = \mathbf{P}'\mathbf{X}$

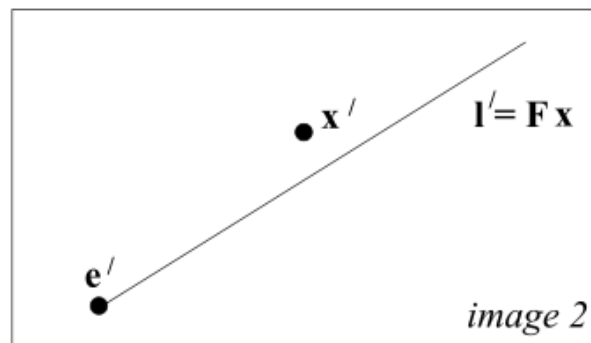
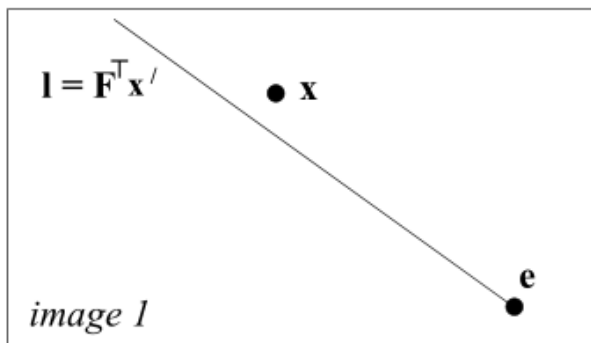


有噪声情况下的三角化

- 由于存在噪声，反投到三维空间上的射线并不会严格相交



优化投影点到对应极线的距离



线性三角化方法

- 给定方程

$$\mathbf{x} = \mathbf{P}\mathbf{X}$$

$$\mathbf{x}' = \mathbf{P}'\mathbf{X}$$

- \mathbf{p}^{iT} 表示 \mathbf{P} 的第 i 行.
- 写成矩阵和向量相乘的形式

$$\begin{bmatrix} x\mathbf{p}^{3T} - \mathbf{p}^{1T} \\ y\mathbf{p}^{3T} - \mathbf{p}^{2T} \\ x'\mathbf{p}'^{3T} - \mathbf{p}'^{1T} \\ y'\mathbf{p}'^{3T} - \mathbf{p}'^{2T} \end{bmatrix} \mathbf{x} = 0$$

- 直接解析求解.
- 没有几何意义 — 不是最优.

优化几何误差

- 目标函数

$$X = \arg \min_X \sum_i ||\pi(\mathbf{P}_i X) - \mathbf{x}_i||^2$$

- 用Levenberg-Marquart算法求解

已知三维，求解相机位姿

- Compute Projection Matrix

$$\mathbf{P}_i = \arg \min_{\mathbf{P}_i} \sum_j \|\pi(\mathbf{P}_i X_j) - \mathbf{x}_{ij}\|^2$$

- Decomposition for Metric Projection Matrix

$$P = K[R \mid t] = [KR \mid Kt] = [M \mid Kt]$$

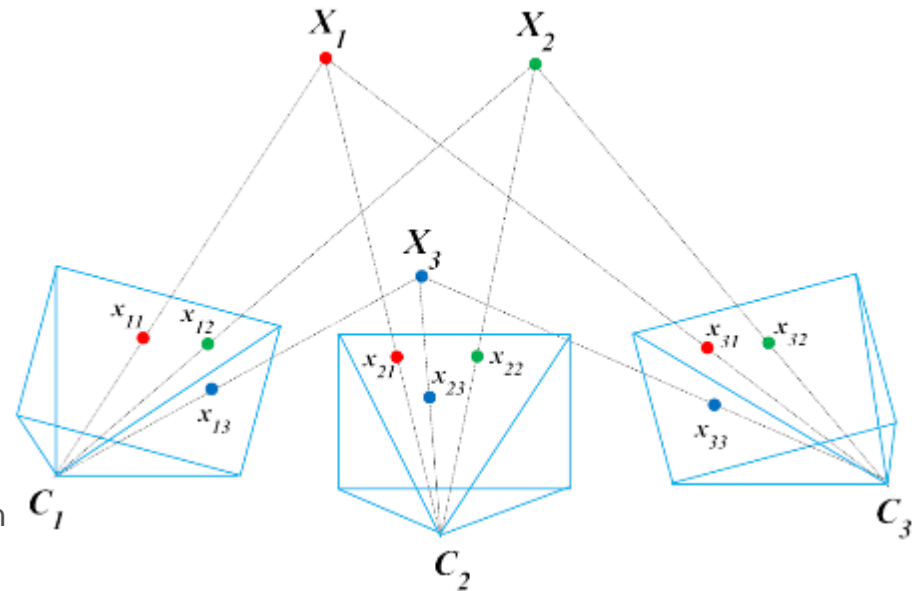
Decompose M into K , R by QR decomposition

$$t = K^{-1}(p_{14}, p_{24}, p_{34})^T$$

Bundle Adjustment

- Definition
 - Refining a visual reconstruction to produce jointly optimal 3D structure and viewing parameter (camera pose and/or calibration) estimates.

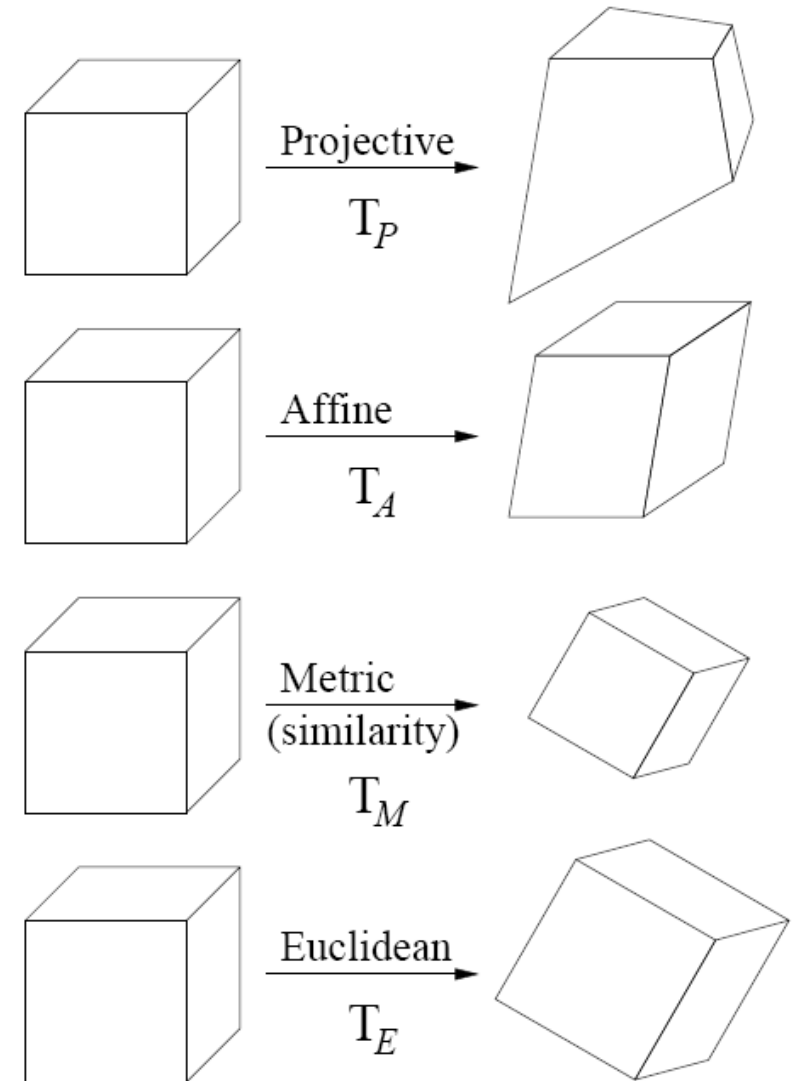
$$\arg \min_{\mathbf{P}_k, X_i} \sum_{k=1}^m \sum_{i=1}^n D(\mathbf{x}_{ki}, \mathbf{P}_k(X_i))^2$$



B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon.
Bundle adjustment - a modern synthesis. In Workshop on Vision
Algorithms, pages 298–372, 1999.

Geometric Ambiguities

ambiguity	DOF	transformation	invariants
projective	15	$\mathbf{T}_P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix}$	cross-ratio
affine	12	$\mathbf{T}_A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix}$	relative distances along direction parallelism <i>plane at infinity</i>
metric	7	$\mathbf{T}_M = \begin{bmatrix} \sigma r_{11} & \sigma r_{12} & \sigma r_{13} & t_x \\ \sigma r_{21} & \sigma r_{22} & \sigma r_{23} & t_y \\ \sigma r_{31} & \sigma r_{32} & \sigma r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$	relative distances angles <i>absolute conic</i>
Euclidean	6	$\mathbf{T}_E = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$	absolute distances



Projective Reconstruction $\xrightarrow{\text{Self-Calibration}}$ Metric Reconstruction

Self-Calibration

- State-of-the-Art References
 - R.I. Hartley and A. Zisserman, [Multiple View Geometry in Computer Vision](#), second ed. Cambridge Univ. Press, 2004.
 - M. Pollefeys, L.J. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch, [Visual Modeling with a Hand-Held Camera](#), Int'l J. Computer Vision, vol. 59, no. 3, pp. 207-232, 2004.
 - G. Zhang, X. Qin, W. Hua, T.-T. Wong, P.-A. Heng, and H. Bao, [Robust Metric Reconstruction from Challenging Video Sequences](#), Proc. IEEE CS Conf. Computer Vision and Pattern Recognition, 2007.

代表性SfM方法

- 增量式SfM

- 采用逐张图片加入处理的方式
- 精度高，求解鲁棒，但速度较慢
- 代表性工作
 - Bundler: <http://grail.cs.washington.edu/rome/>
 - VisualSFM: <http://ccwu.me/vsfm/>
 - COLMAP: <http://demuc.de/colmap/>
 - ACTS: <http://www.zjucvg.net/acts/acts.html>

- 层次式SfM

- 先求解局部地图，再进行融合和补充得到完整的重建
- 显著提高重建的效率
- 代表性工作
 - ENFT-SFM: <https://github.com/zju3dv/ENFT-SfM>

代表性SfM方法

- 全局式SfM
 - 直接求解全局的图像外参，然后通过少量的集束调整完成优化
 - 高效，但是容易受到错误匹配的影响
 - 代表性工作
 - OpenMVG: <https://github.com/openMVG/openMV>
- 混合式SfM
 - 采用全局方法估计图像旋转，然后增量式的求解图像位置
 - 大幅度减少重建时间，错误的匹配关系可以及时修正
 - 代表性工作
 - HSfM: Hybrid Structure-from-Motion(Cui et al.,2017)
- 语义SfM
 - 通过语义信息进行联合优化
 - 三维地图包含语义信息
 - 代表工作
 - Semantic structure from motion (Bao et al.,2011)

谢谢！