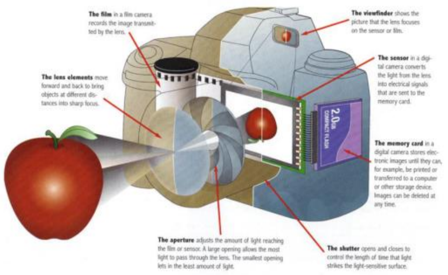
相机标定与稀疏重建

申抒含
中国科学院自动化研究所
模式识别国家重点实验室

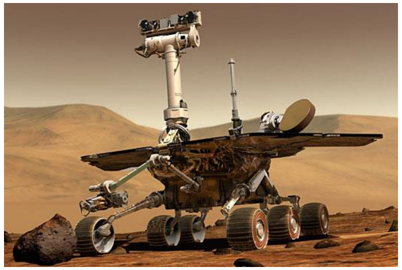
Robot Vision Group
National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences

三维计算机视觉

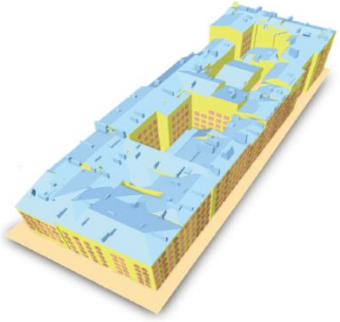
6、相机模型与多视几何



7、相机标定与稀疏重建

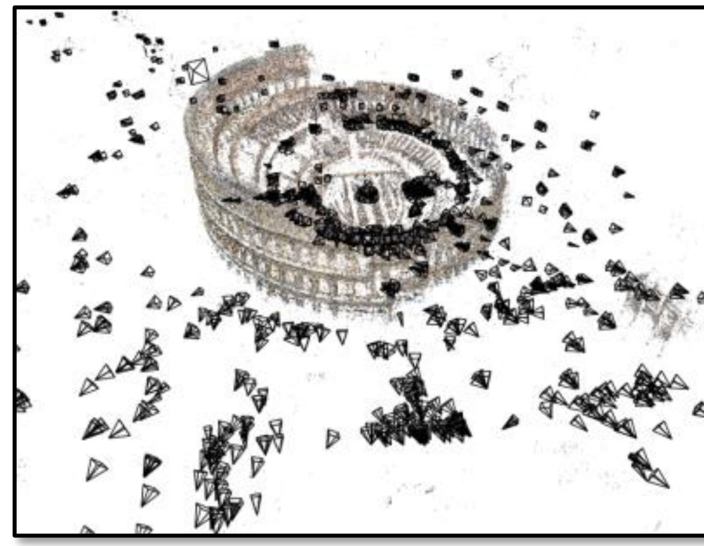


8、立体视觉与三维建模



相机标定与稀疏重建

多视角图像



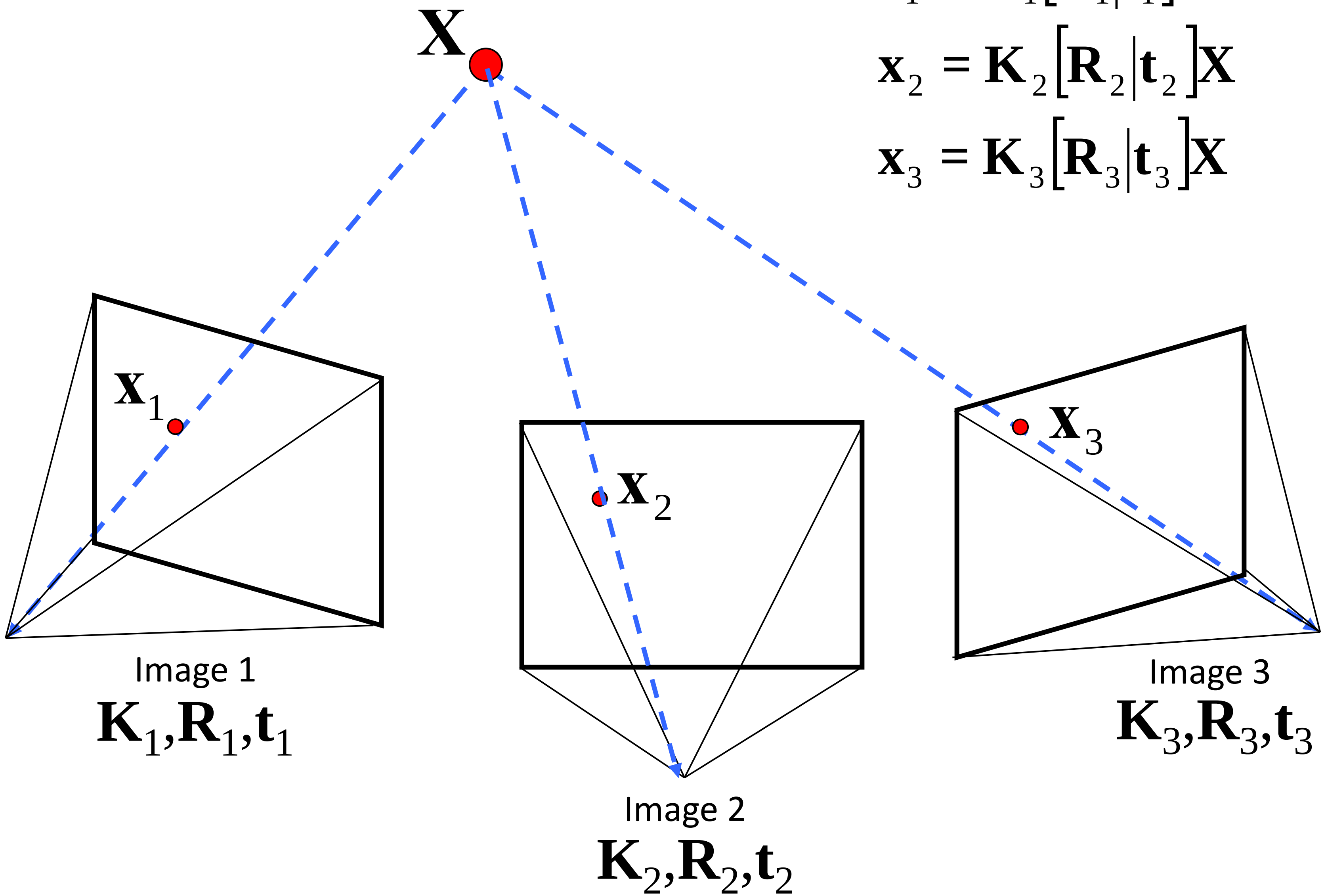
相机位姿+场景结构

多视几何

$$\mathbf{x}_1 = \mathbf{K}_1\left[\mathbf{R}_1 \middle| \mathbf{t}_1\right]\mathbf{X}$$

$$\mathbf{x}_2 = \mathbf{K}_2\left[\mathbf{R}_2 \middle| \mathbf{t}_2\right]\mathbf{X}$$

$$\mathbf{x}_3 = \mathbf{K}_3\left[\mathbf{R}_3 \middle| \mathbf{t}_3\right]\mathbf{X}$$



相机标定与稀疏重建

$$\mathbf{x}_1 = \mathbf{K}_1 \left[\mathbf{R}_1 \middle| \mathbf{t}_1 \right] \mathbf{X}$$
$$\vdots$$
$$\mathbf{x}_n = \mathbf{K}_n \left[\mathbf{R}_n \middle| \mathbf{t}_n \right] \mathbf{X}$$

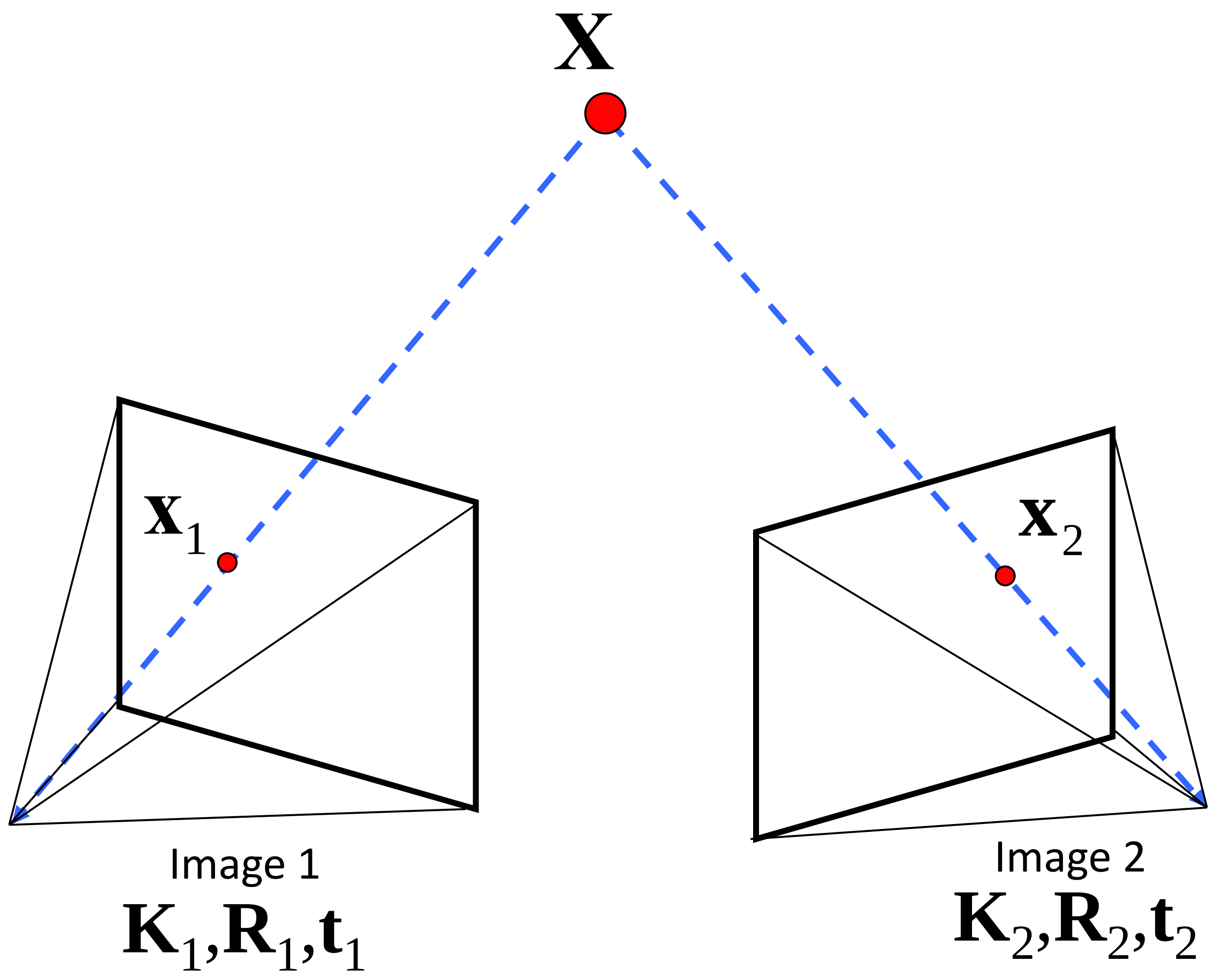
- 已知**x**、**K**、**R**、**t**，求**X**： 三角化（Triangulation）
- 已知**x**、**X**，求**K**、**R**、**t**： 相机标定（Camera Calibration）
- 已知**x**，求**K**、**R**、**t**、**X**： 稀疏重建（Sparse Reconstruction）

  Structure from Motion (SfM)

  Structure and Motion Estimation

三角化（Triangulation）

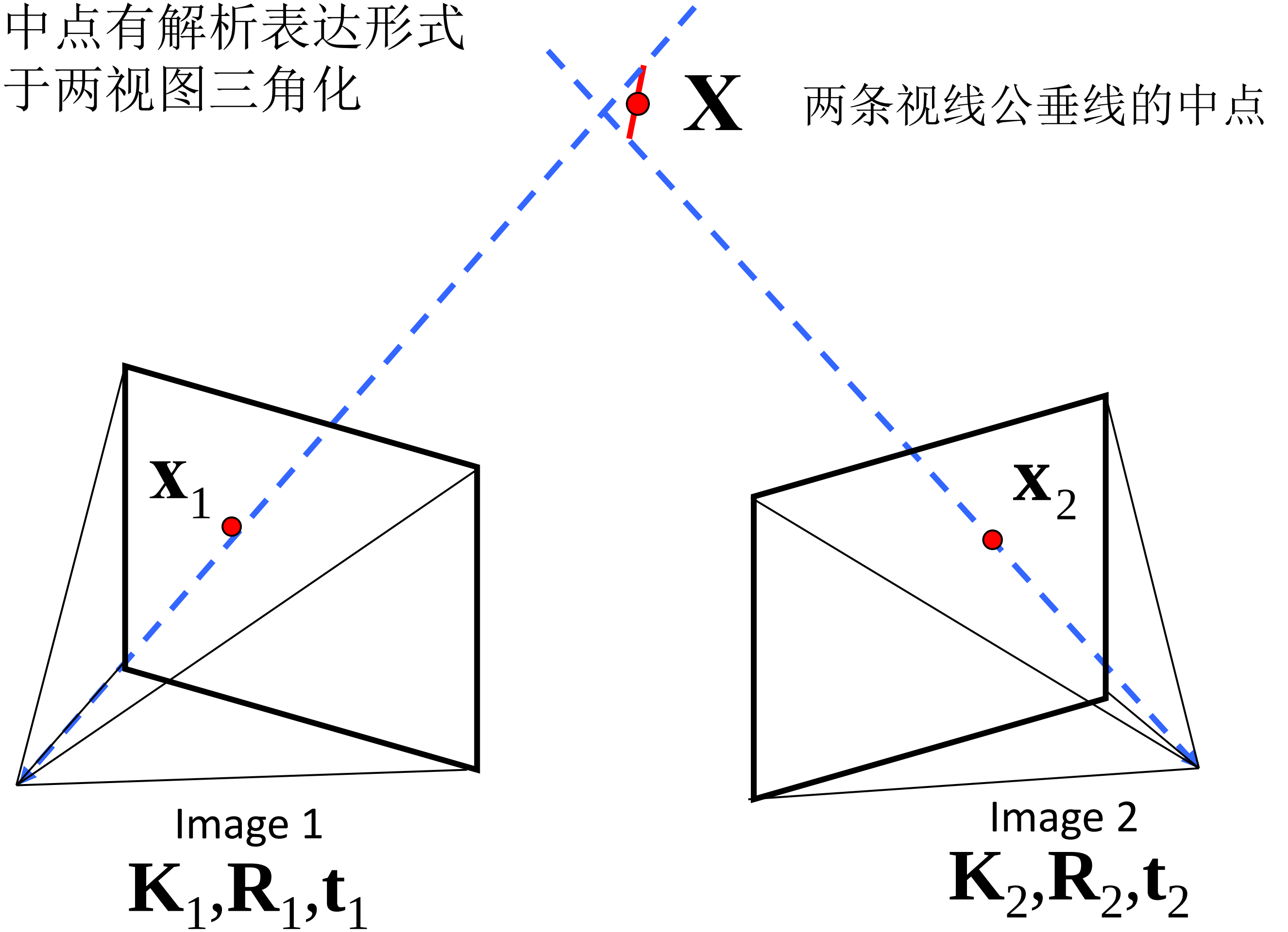
- 已知**x**、**K**、**R**、**t**，求**X**



三角化（Triangulation）

- 已知**x**、**K**、**R**、**t**，求**X**

—公垂线中点有解析表达形式
—只是用于两视图三角化



三角化（Triangulation）

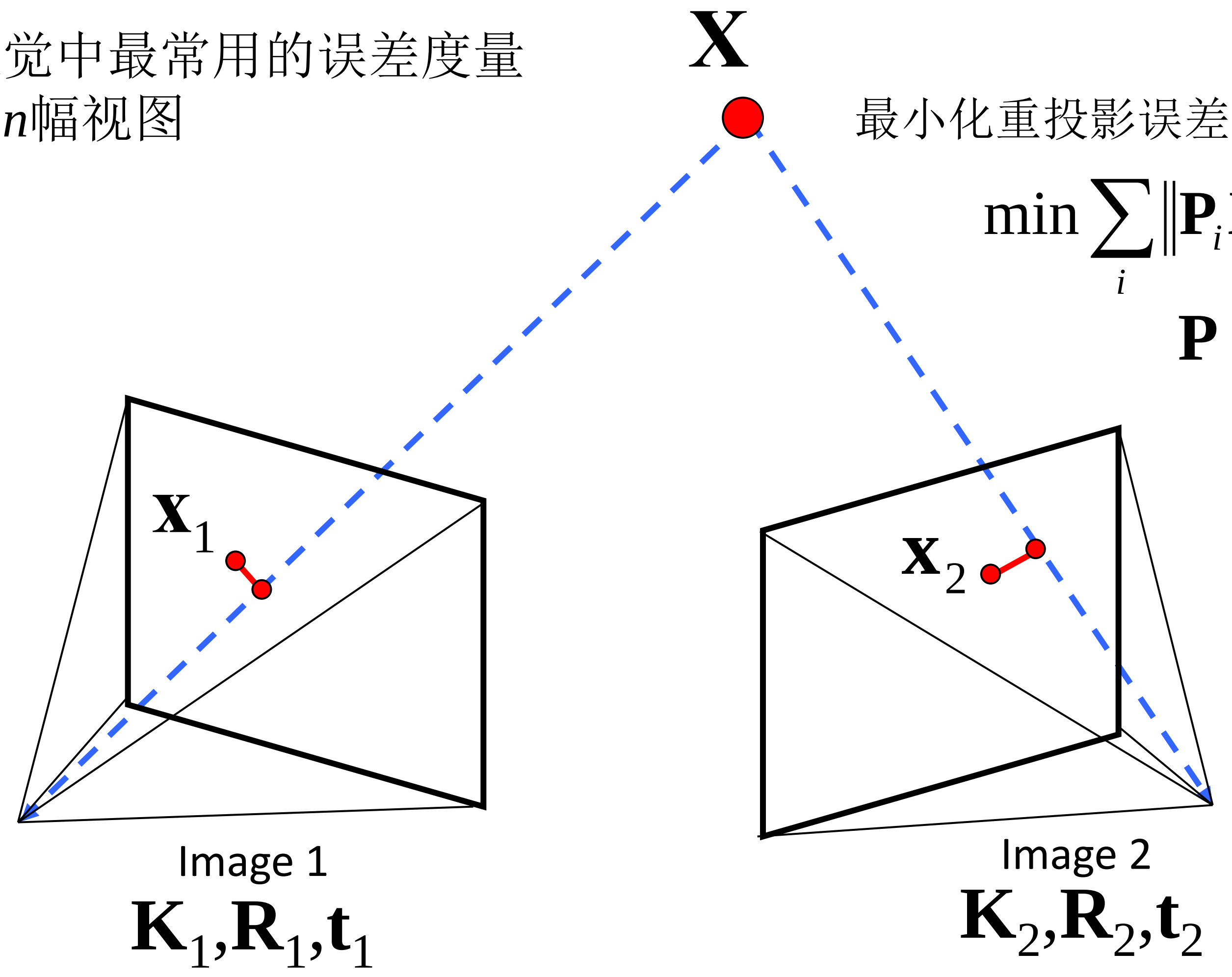
- 已知**x**、**K**、**R**、**t**，求**X**

—三维视觉中最常用的误差度量
—适用于*n*幅视图

最小化重投影误差平方和

$$\min \sum_i \left\| \mathbf{P}_i \mathbf{X} - \mathbf{x}_i \right\|^2$$

$$\mathbf{P} = \mathbf{K}\left[\mathbf{R} \middle| \mathbf{t}\right]$$



三角化（Triangulation）

- 已知**x**、**K**、**R**、**t**，求**X**

重投影误差平方和最小化的求解 $\min \sum_i \left\| \mathbf{P}_i \mathbf{X} - \mathbf{x}_i \right\|^2$

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

令: $\mathbf{P} = \begin{bmatrix} \mathrm{P}_1 \\ \mathrm{P}_2 \\ \mathrm{P}_3 \end{bmatrix}$ **P**为3×4矩阵，P_i为1×4向量

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \mathrm{P}_1 \\ \mathrm{P}_2 \\ \mathrm{P}_3 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \Rightarrow \begin{aligned} x &= \frac{\mathrm{P}_1[X \quad Y \quad Z \quad 1]^T}{\mathrm{P}_3[X \quad Y \quad Z \quad 1]^T} \\ y &= \frac{\mathrm{P}_2[X \quad Y \quad Z \quad 1]^T}{\mathrm{P}_3[X \quad Y \quad Z \quad 1]^T} \end{aligned}$$

三角化（Triangulation）

- 已知**x**、**K**、**R**、**t**，求**X**

重投影误差平方和最小化的求解 $\min \sum_i \left\| \mathbf{P}_i \mathbf{X} - \mathbf{x}_i \right\|^2$

$$x = \frac{\mathrm{P}_1 \begin{bmatrix} X & Y & Z & 1 \end{bmatrix}^T}{\mathrm{P}_3 \begin{bmatrix} X & Y & Z & 1 \end{bmatrix}^T} \qquad x\mathrm{P}_3 \begin{bmatrix} X & Y & Z & 1 \end{bmatrix}^T = \mathrm{P}_1 \begin{bmatrix} X & Y & Z & 1 \end{bmatrix}^T$$

$$y = \frac{\mathrm{P}_2 \begin{bmatrix} X & Y & Z & 1 \end{bmatrix}^T}{\mathrm{P}_3 \begin{bmatrix} X & Y & Z & 1 \end{bmatrix}^T} \qquad y\mathrm{P}_3 \begin{bmatrix} X & Y & Z & 1 \end{bmatrix}^T = \mathrm{P}_2 \begin{bmatrix} X & Y & Z & 1 \end{bmatrix}^T$$

n幅图像可得关于**X**的$2n$个线性方程：

$$\mathbf{A}_{2n\times 3} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \mathrm{b}_{2n\times 1} \quad \Rightarrow \quad \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \mathbf{A}^+ \mathrm{b}_{2n\times 1} \qquad \mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$$

三角化（Triangulation）

- 已知**x**、**K**、**R**、**t**，求**X**

重投影误差平方和最小化的求解 $\min \sum_i \left\| \mathbf{P}_i \mathbf{X} - \mathbf{x}_i \right\|^2$

以线性解 $\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \mathbf{A}^+ \mathrm{b}_{2n\times 1}$ 为初始值，迭代求解非线性优化问题

几何误差

代数误差

基本思路：以代数误差解为初值，迭代求解几何误差解

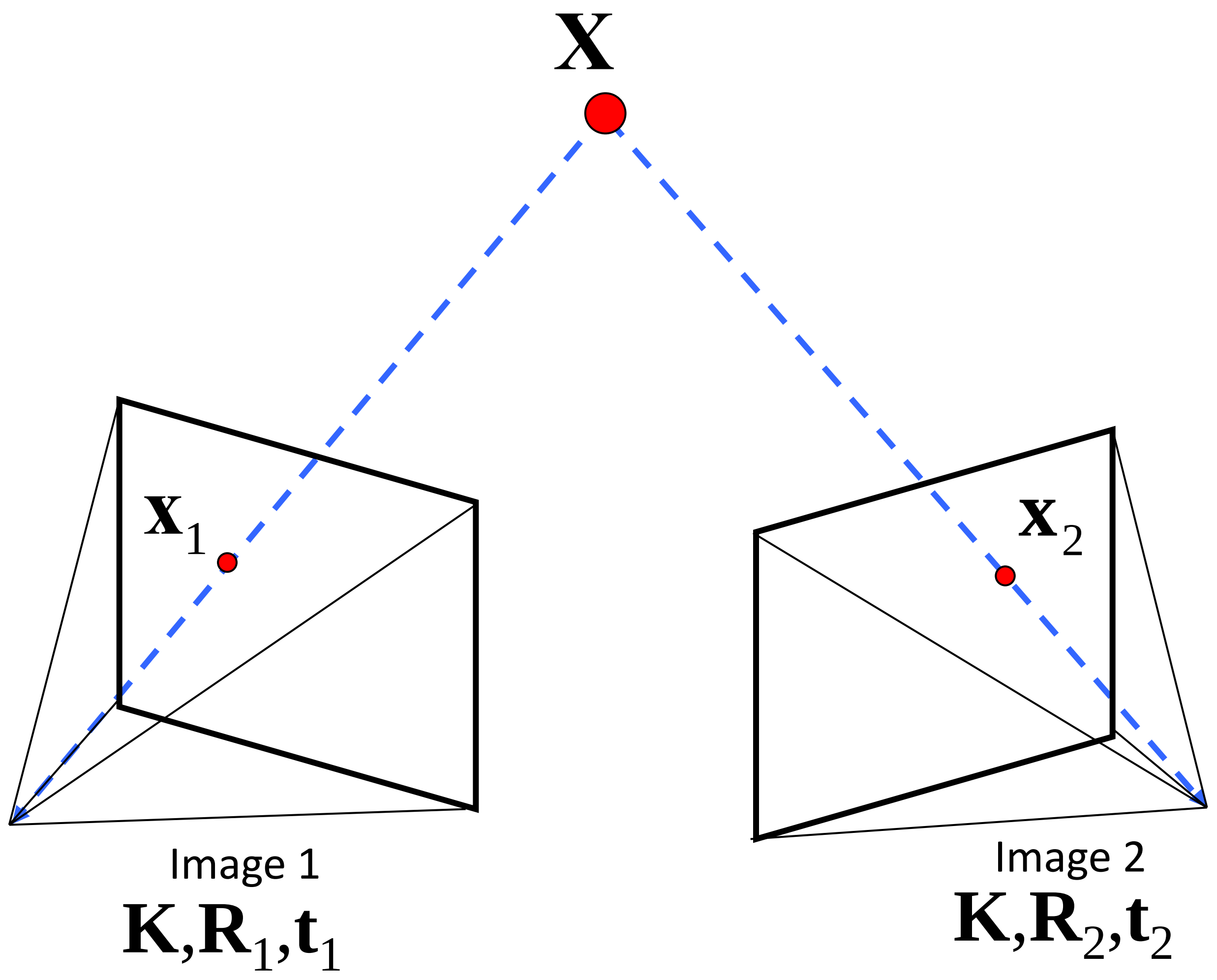
三角化（Triangulation）

- Triangulation的应用

DJI PHANTOM 4

相机标定（Camera Calibration）

- 已知**x**、**X**，求**K**、**R**、**t**



相机标定（Camera Calibration）

- 已知**x**、**X**，求**K**、**R**、**t**

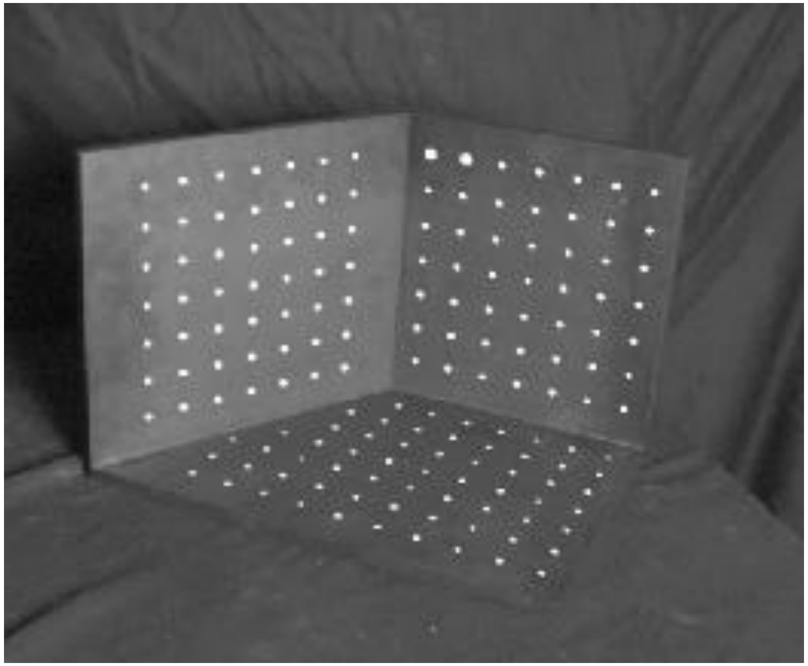
$$\mathbf{x} = \mathbf{K}\left[\mathbf{R} \middle| \mathbf{t}\right]\mathbf{X}$$

相机内参数　　　　相机外参数（相机位姿）

- 需要知道一组2D—3D对应，并知道对应点的2D与3D坐标

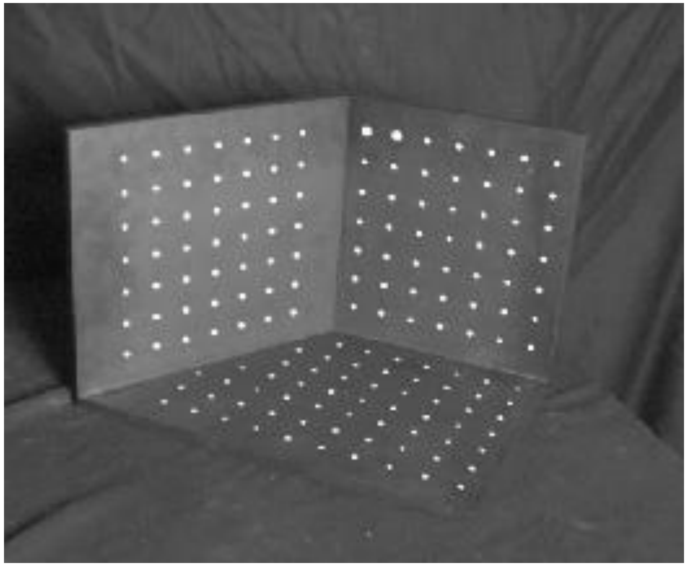
相机标定（Camera Calibration）

- 通常情况下，需要一个人工制作的标定物



要求标定物结构已知，标定物上特征点易于提取

相机标定（Camera Calibration）



$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}$$

相机标定（Camera Calibration）

- 一组2D—3D对应点提供关于**P**的两个线性方程

$$\begin{bmatrix} X_i & Y_i & Z_i & 1 & 0 & 0 & 0 & 0 & -x_iX_i & -x_iY_i & -x_iZ_i & -x_i \\ 0 & 0 & 0 & 0 & X_i & Y_i & Z_i & 1 & -y_iX & -y_iY & -y_iZ_i & -y_i \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ p_{13} \\ p_{14} \\ p_{21} \\ p_{22} \\ p_{23} \\ p_{24} \\ p_{31} \\ p_{32} \\ p_{33} \\ p_{34} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

相机标定（Camera Calibration）

- n组2D—3D对应点提供关于**P**的2 n个线性方程

$$
\begin{bmatrix}
X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -x_1X_1 & -x_1Y_1 & -x_1Z_1 & -x_1 \\
0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -y_1X & -y_1Y & -y_1Z_1 & -y_1 \\
\vdots & & & & & & & & & & & \\
X_n & Y_n & Z_n & 1 & 0 & 0 & 0 & 0 & -x_nX_n & -x_nY_n & -x_nZ_n & -x_n \\
0 & 0 & 0 & 0 & X_n & Y_n & Z_n & 1 & -y_nX & -y_nY & -y_nZ_n & -y_n
\end{bmatrix}
\begin{bmatrix}
p_{11} \\ p_{12} \\ p_{13} \\ p_{14} \\ p_{21} \\ p_{22} \\ p_{23} \\ p_{24} \\ p_{31} \\ p_{32} \\ p_{33} \\ p_{34}
\end{bmatrix}
=
\begin{bmatrix} 0 \\ 0 \end{bmatrix}
$$

$$\mathbf{A}\,\mathrm{P} = 0$$

Direct Linear Transformation (DLT)

相机标定（Camera Calibration）

- DLT求解相机标定的优缺点：

- 优点：
  —线性求解

- 缺点：
  —需要从**P**中分解相机内外参数**K**、**R**、**t**
  —不包含相机畸变参数
  —无法添加其他约束（如已知相机焦距）
  —最小化代数误差（无几何意义）

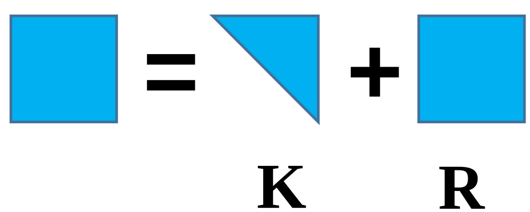
相机标定（Camera Calibration）

从**P**中分解相机内外参数**K**、**R**、**t**:

$$\mathbf{P}=\mathbf{K}\left[\mathbf{R}\middle|\mathbf{t}\right]=[\mathbf{KR}|\mathbf{Kt}]$$

$$\mathbf{K}=\begin{bmatrix} f_x & s & p_x \\ 0 & f_y & p_y \\ 0 & 0 & 1 \end{bmatrix}$$ 是上三角阵，**R**是正交矩阵

对**P**的前三列进行RQ分解:



$\mathbf{K}^{-1}$乘以**P**的第四列得到**t**

相机标定（Camera Calibration）

- 基本思路：以DLT代数误差解为初值，迭代求解几何误差解

$$\min \sum_i \left\| \mathbf{P}\mathbf{X}_i - \mathbf{x}_i \right\|^2$$

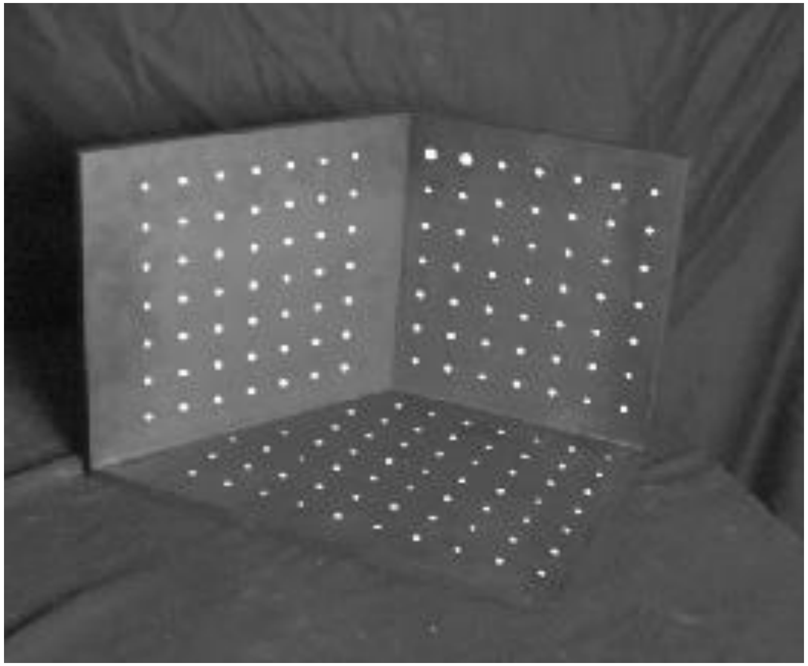
- 区别相机标定与三角化求解时的差异：

三角化：$\min \sum_i \left\| \mathbf{P}_i \mathbf{X} - \mathbf{x}_i \right\|^2$ $\mathbf{x}$、$\mathbf{P}$已知，求$\mathbf{X}$

相机标定：$\min \sum_i \left\| \mathbf{P}\mathbf{X}_i - \mathbf{x}_i \right\|^2$ $\mathbf{x}$、$\mathbf{X}$已知，求$\mathbf{P}$

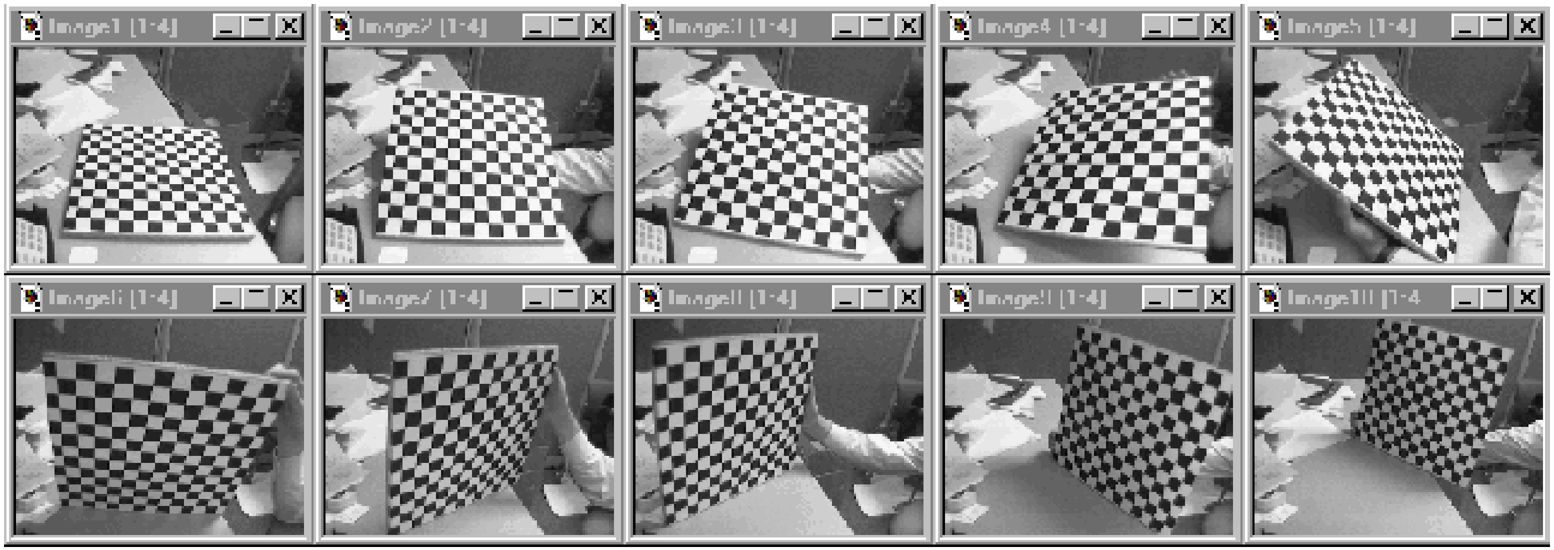
相机标定（Camera Calibration）

- 使用三维标定物的优缺点：
- 优点
  - 标定精度高
  - 通过一幅图像即可标定
- 缺点
  - 需要高精度的三维标定块



相机标定（Camera Calibration）

- 更实用的标定方案：使用平面标定板

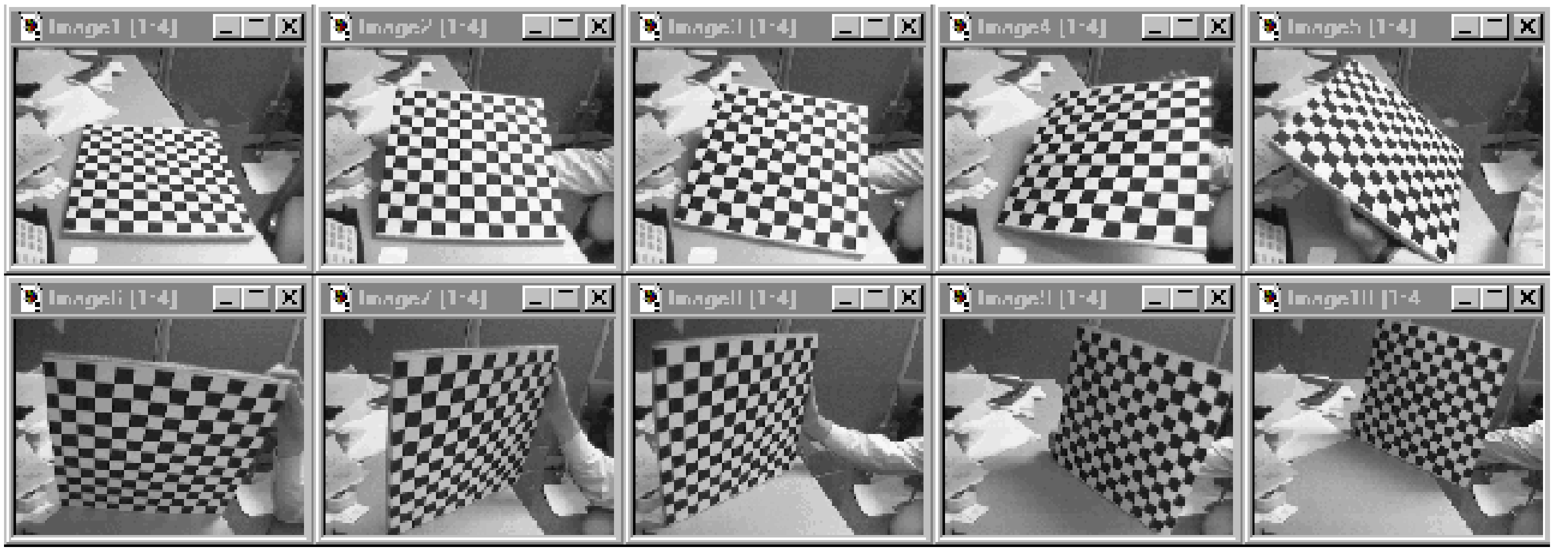


- 优点
  - —容易制作（打印一张黑白棋盘格、一块足够平的木板）
  - —标定工具箱成熟（Matlab、OpenCV）

Zhengyou Zhang: A Flexible New Technique for Camera Calibration. IEEE T-PAMI), 2000.

相机标定（Camera Calibration）

- 更实用的标定方案：使用平面标定板

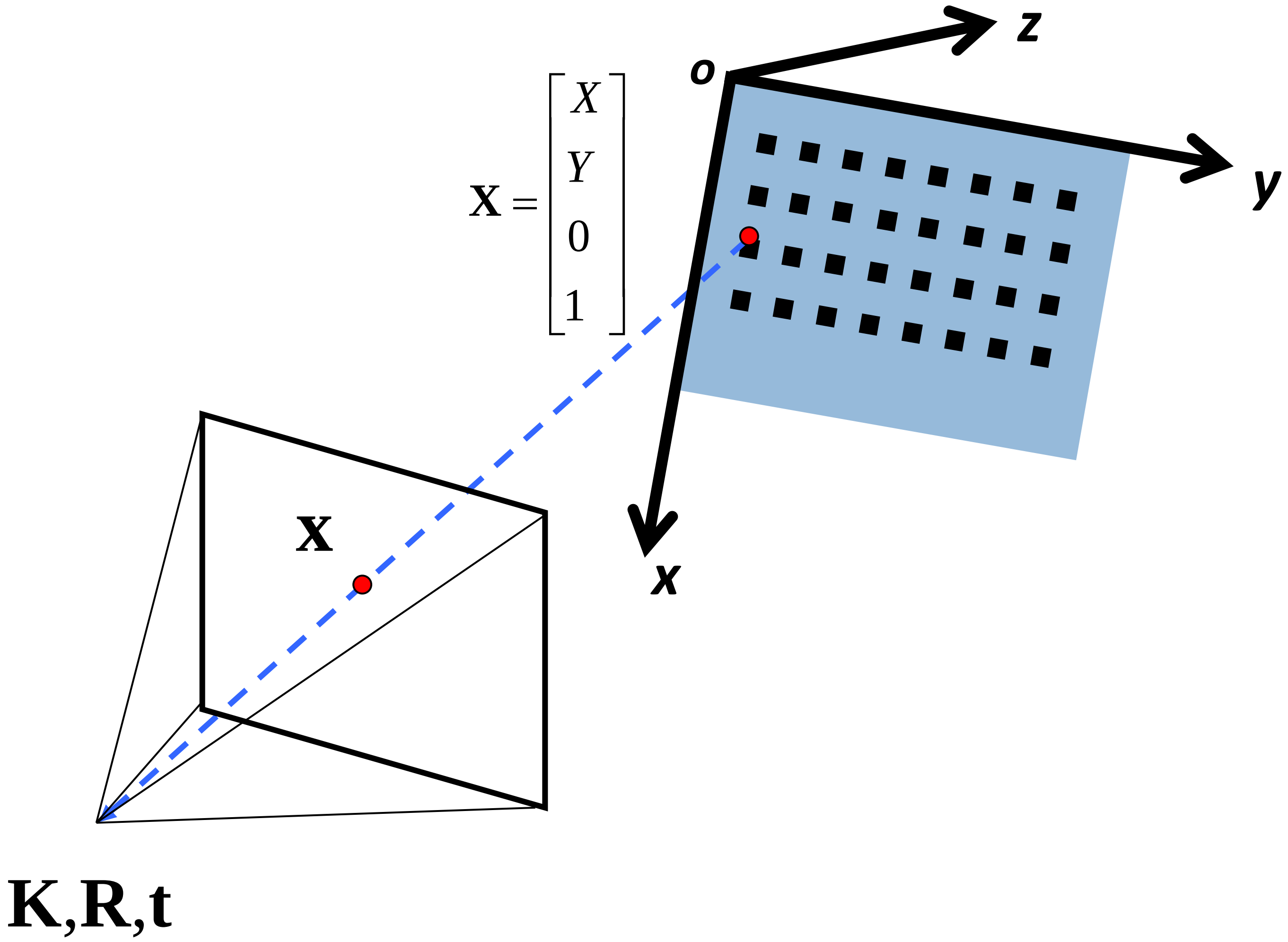


- 缺点
  - —标定精度不如三维标定物

Zhengyou Zhang: A Flexible New Technique for Camera Calibration. IEEE T-PAMI), 2000.

相机标定（Camera Calibration）

- 更实用的标定方案：使用平面标定板



相机标定（Camera Calibration）

- 平面标定板

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{K}\left[\mathbf{R} \middle| \mathbf{t}\right]\begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = \mathbf{K}\left[\mathrm{r}_1 \quad \mathrm{r}_2 \quad \mathrm{r}_3 \middle| \mathbf{t}\right]\begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = \mathbf{K}\underbrace{\left[\mathrm{r}_1 \quad \mathrm{r}_2 \middle| \mathbf{t}\right]}_{\mathbf{H}_{3\times 3}}\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

- **H**称为单应，有8个自由度，可通过4组对应点DLT线性求解
- 令 $\mathbf{H} = [\mathrm{h}_1 \quad \mathrm{h}_2 \quad \mathrm{h}_3] = \mathbf{K}\left[\mathrm{r}_1 \quad \mathrm{r}_2 \middle| \mathbf{t}\right]$

相机标定（Camera Calibration）

- 从**H**中求解**K**：

$$\mathbf{H} = [h_1 \quad h_2 \quad h_3] = \mathbf{K}[r_1 \quad r_2 \mid \mathbf{t}] \Rightarrow \mathbf{K}^{-1}[h_1 \quad h_2 \quad h_3] = [r_1 \quad r_2 \mid \mathbf{t}]$$

由正交矩阵的性质：

$$r_1^T r_2 = 0 \qquad \|r_1\| = \|r_2\| = 1$$

可得：

$$h_1^T \mathbf{K}^{-T}\mathbf{K}^{-1} h_2 = 0$$

$$h_1^T \mathbf{K}^{-T}\mathbf{K}^{-1} h_1 = h_2^T \mathbf{K}^{-T}\mathbf{K}^{-1} h_2$$

每幅图像可以获得2个内参数约束方程，对于5参数**K**，当图像数目≥3时，可以线性唯一求解出**K**

相机标定（Camera Calibration）

- 求解**R**、**t**：

$$\mathbf{H} = \mathbf{K}\left[r_1 \quad r_2 \,|\, \mathbf{t} \right]$$

根据前面求解出的**K**：

$$\left[r_1 \quad r_2 \,|\, \mathbf{t} \right] = \mathbf{K}^{-1}\mathbf{H}$$

可得：

$$r_3 = r_1 \times r_2 \qquad \mathbf{R} = \left[r_1 \quad r_2 \quad r_3 \right]$$

至此，求解出相机内外参数**K**、**R**、**t**

相机标定（Camera Calibration）

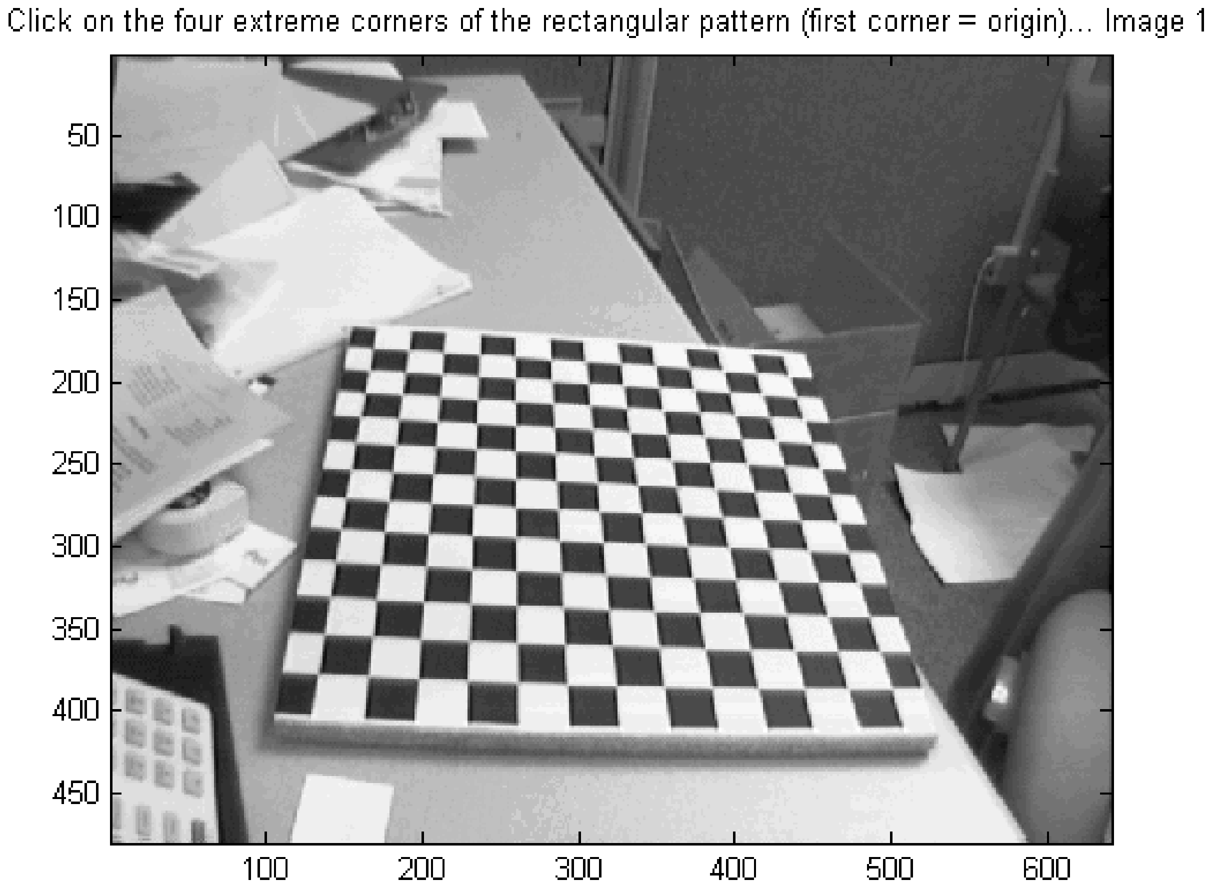
- 以上述线性求解的**K**、**R**、**t**为初始值（代数误差最小化解），迭代求解重投影误差最小化问题：

$$\min \sum_i \left\| \mathbf{PX}_i - \mathbf{x}_i \right\|^2$$

得到几何误差最小化意义下的**K**、**R**、**t**

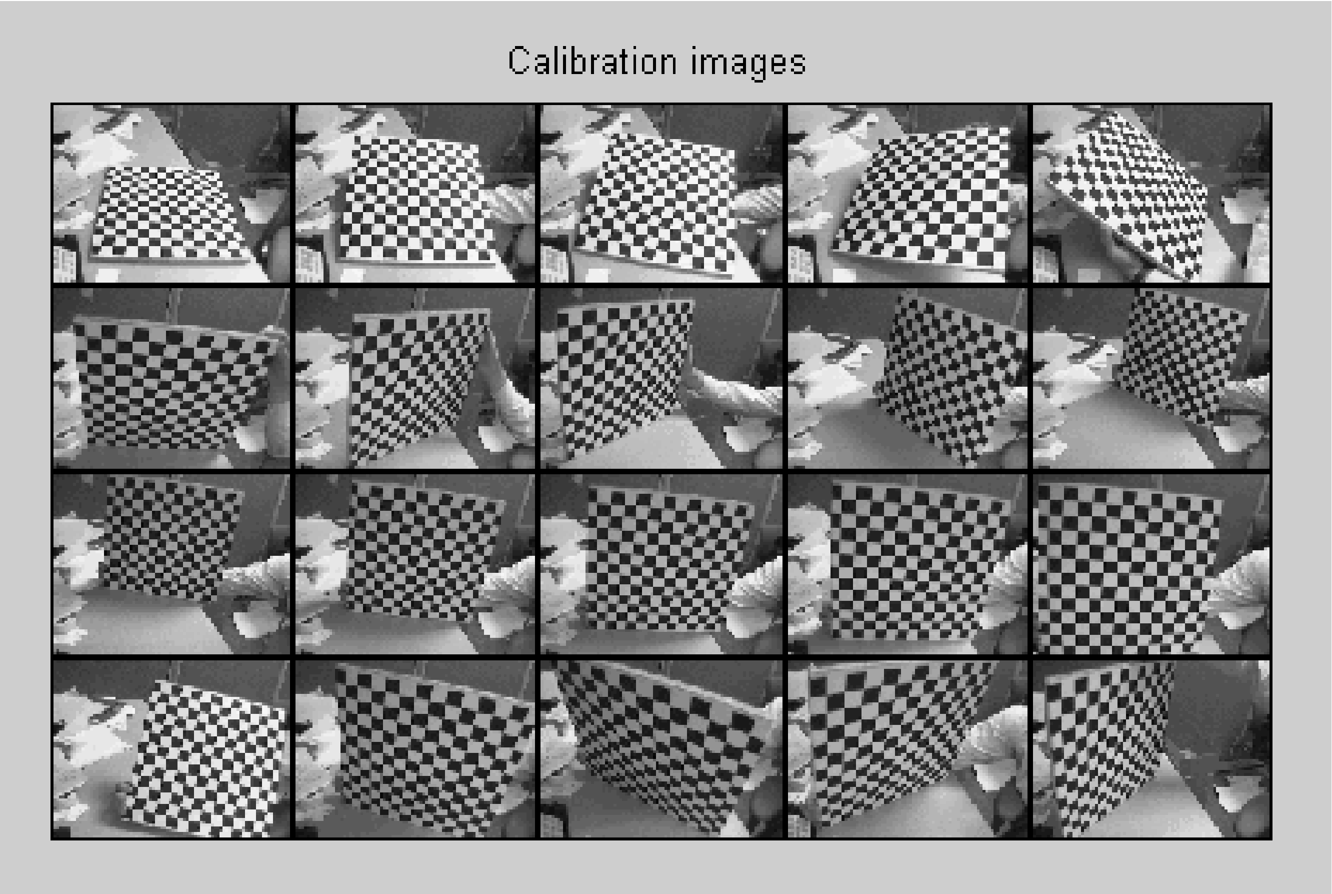
相机标定（Camera Calibration）

- 标定流程1：打印一张模板并贴在一个平面上



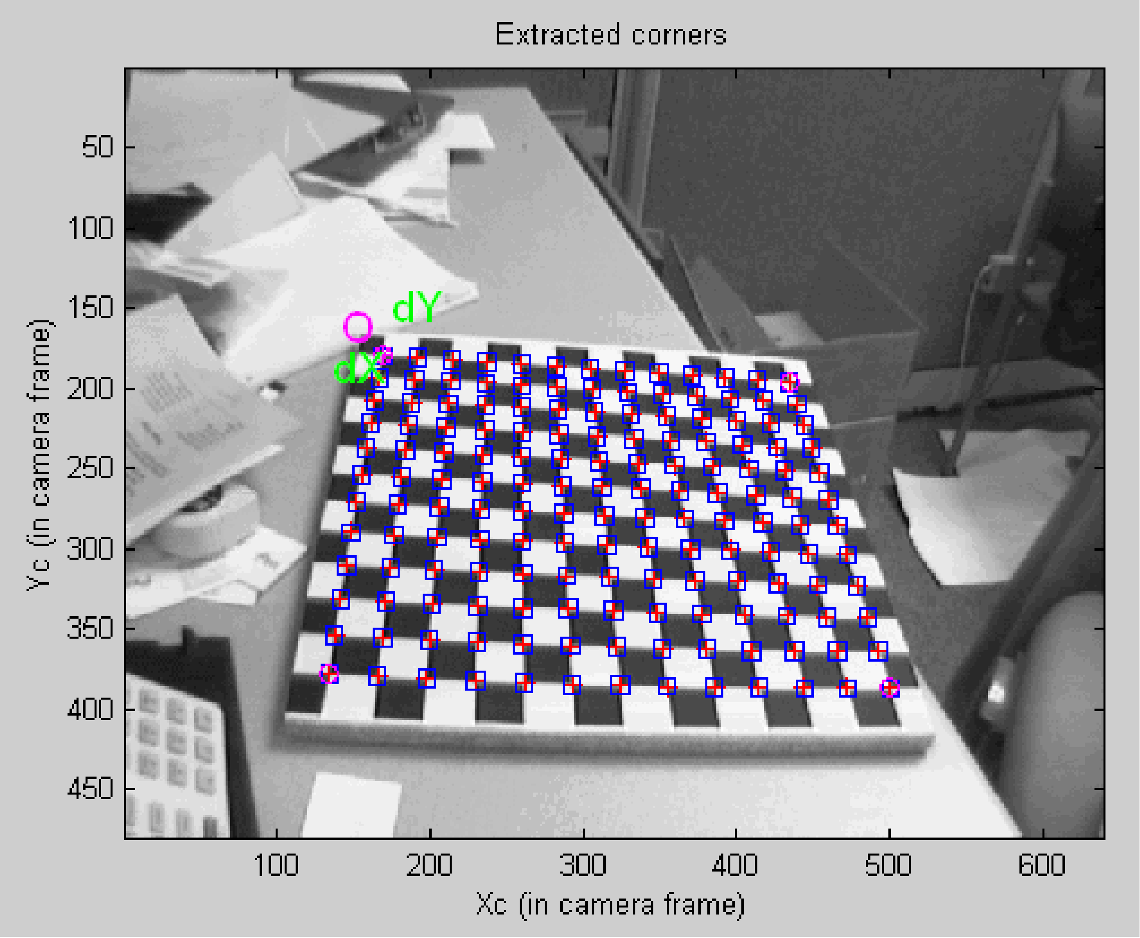
相机标定（Camera Calibration）

- 标定流程2：从不同角度拍摄若干张模板图像



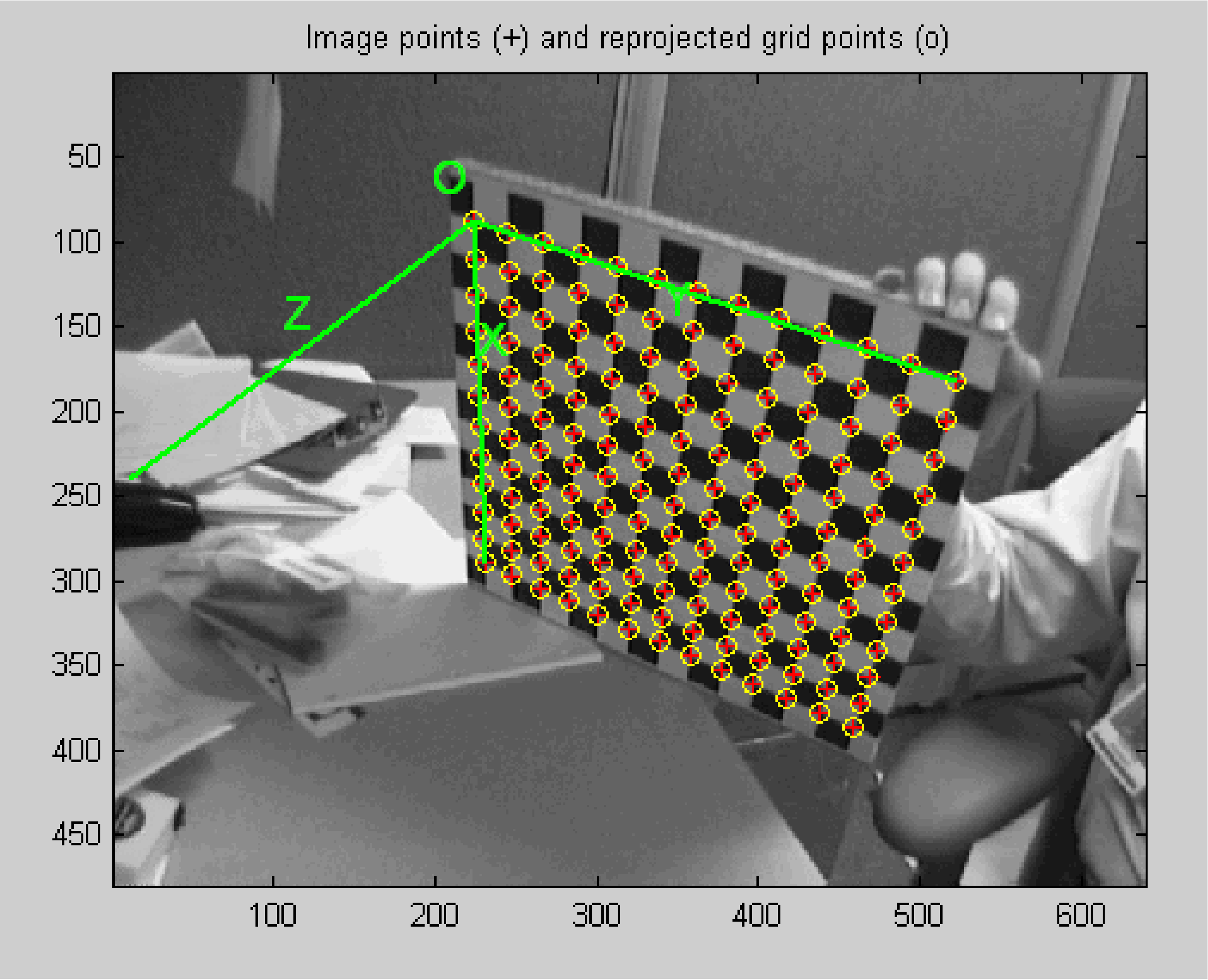
相机标定（Camera Calibration）

- 标定流程3：检测图像中的特征点



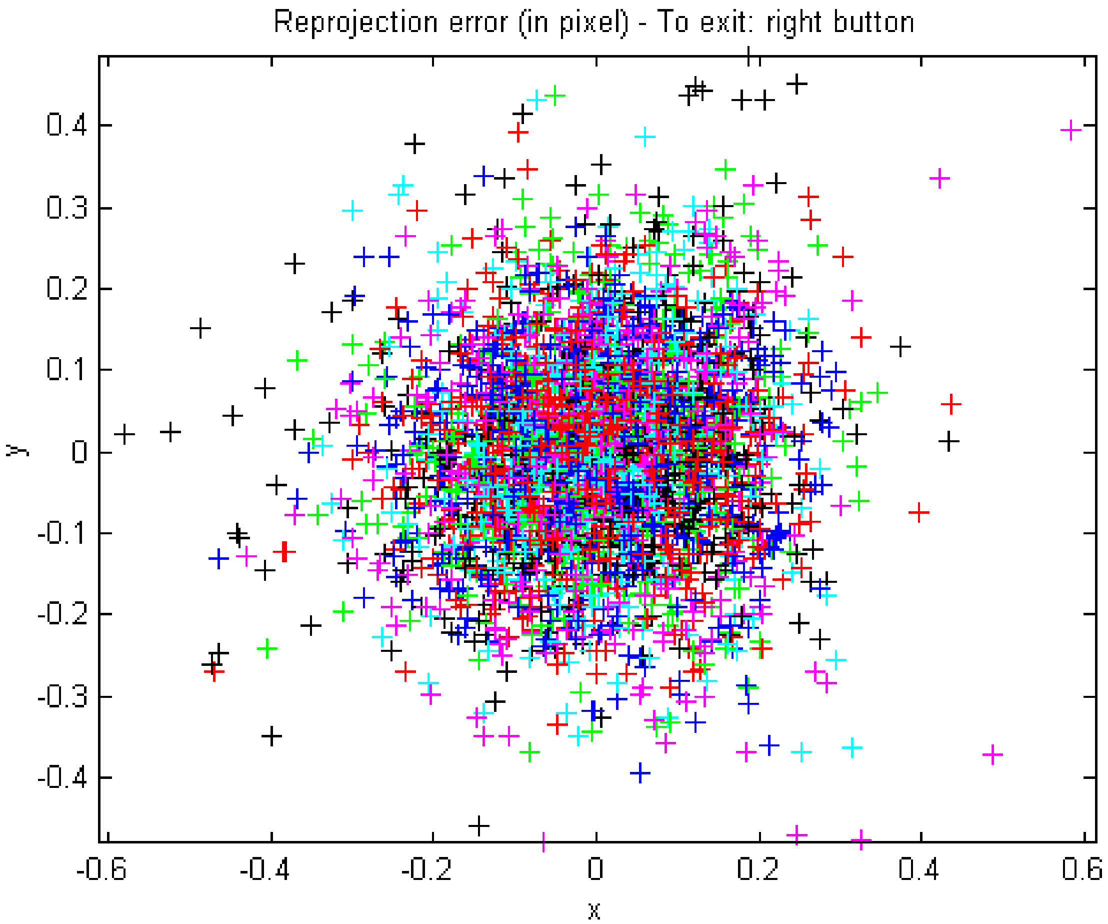
相机标定（Camera Calibration）

- 标定流程4：求解相机内外参数



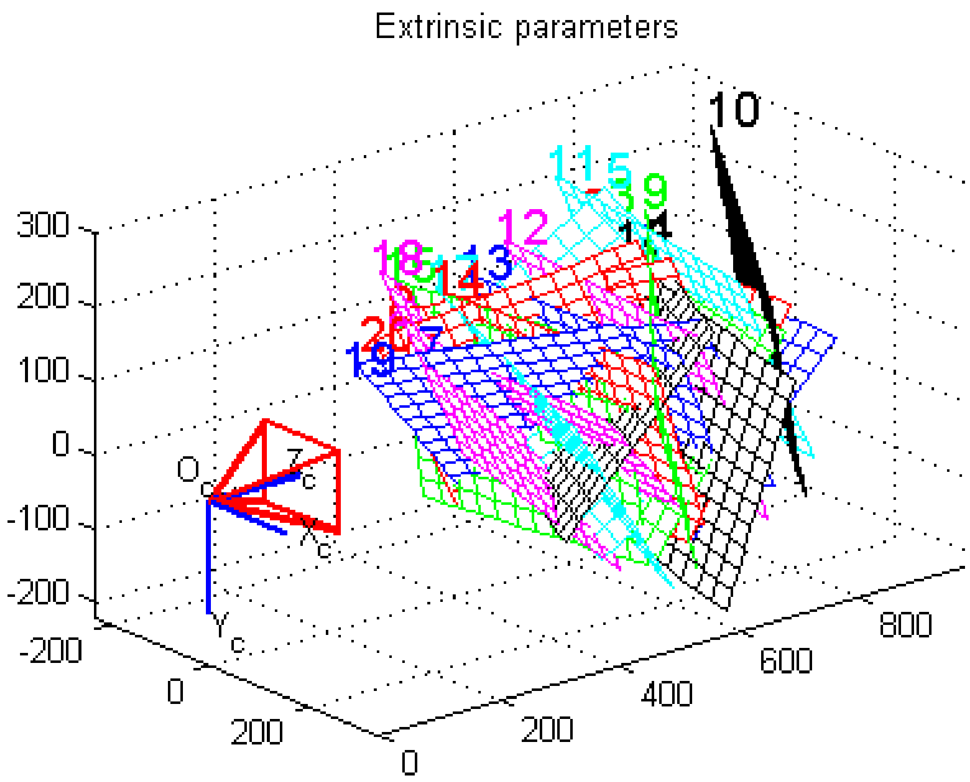
相机标定（Camera Calibration）

- 标定流程5：分析重投影误差



相机标定（Camera Calibration）

- 标定流程6：输出标定结果

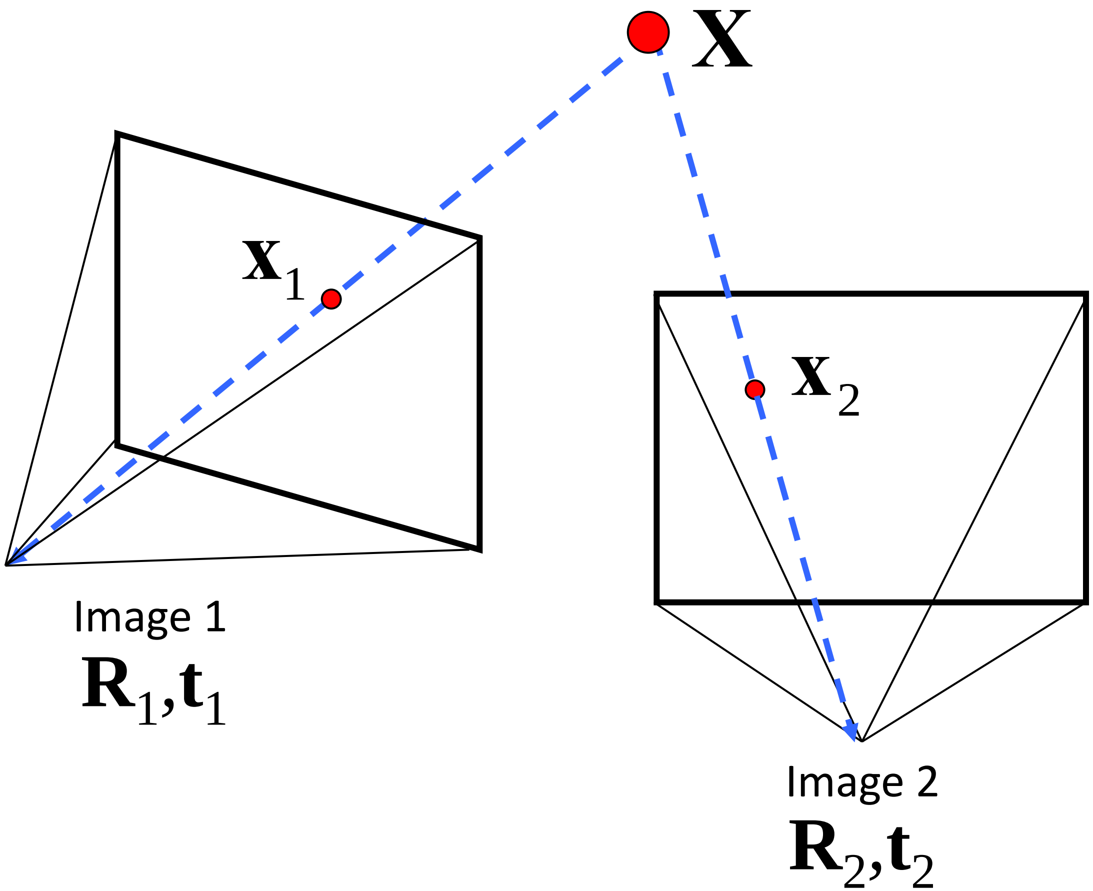


相机标定（Camera Calibration）

- 机器换人

稀疏重建（Sparse Reconstruction）

- 已知**x**，求**K**、**R**、**t** 、**X**
- 也叫Structure from Motion或Structure and Motion Estimation



$$\mathbf{x}_1^T \mathbf{F} \mathbf{x}_2 = 0$$

8点法求F（上节课）

稀疏重建（Sparse Reconstruction）

- 基本矩阵**F**和本质矩阵**E**

$$\mathbf{x}_1^T \mathbf{F} \mathbf{x}_2 = 0$$

$$\Downarrow$$

$$\mathbf{x}^T{}_1 \mathbf{K}_1^{-T} \mathbf{R}[\mathbf{t}]_\times \mathbf{K}_2^{-1} \mathbf{x}_2 = 0$$

- 当内参数矩阵$\mathbf{K}_1$和$\mathbf{K}_2$已知时，$\mathbf{E}=\mathbf{R}[\mathbf{t}]_\times$称为本质矩阵

$$\tilde{\mathbf{x}}^T{}_1 \mathbf{E} \tilde{\mathbf{x}}_2 = 0$$

$$\tilde{\mathbf{x}}_1 = \mathbf{K}_1^{-1} \mathbf{x}_1 \qquad \tilde{\mathbf{x}}_2 = \mathbf{K}_2^{-1} \mathbf{x}_2$$

稀疏重建（Sparse Reconstruction）

- 5点法求解本质矩阵**E**

$$\tilde{\mathbf{x}}^T{}_1\mathbf{E}\tilde{\mathbf{x}}_2 = 0$$

Nister, “An efficient solution to the 5-point relative pose problem,” IEEE T-PAMI 2004

稀疏重建（Sparse Reconstruction）

- 通过SVD分解从**E**中分解**R**和**t**（假设camera 1为$\mathbf{R}_1$=**I**, $\mathbf{t}_1$=0）：

$$\mathbf{E} = \mathbf{U}\,\mathrm{diag}(1,1,0)\mathbf{V}^T$$

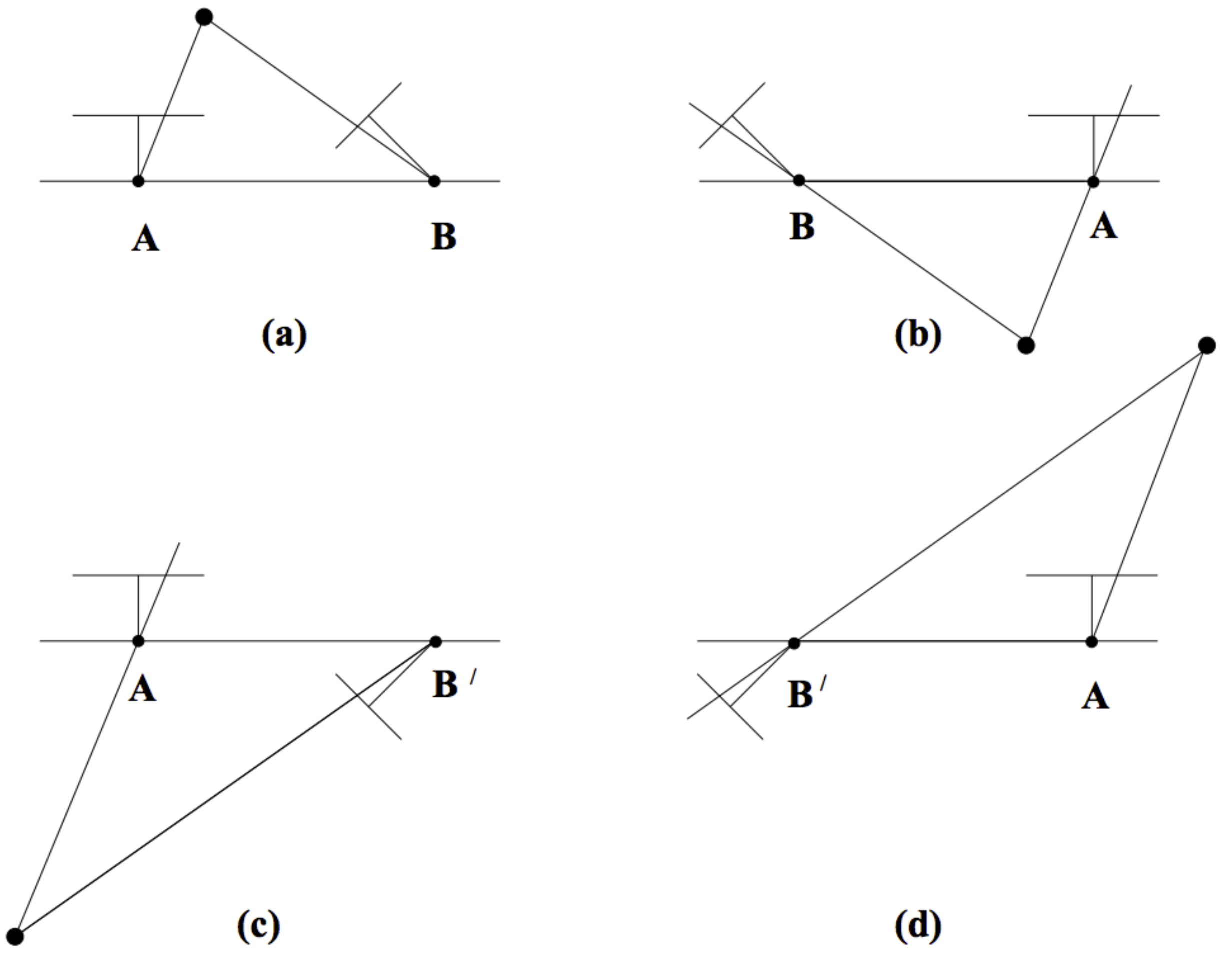
则camera 2有四组解：

$$[\mathbf{R}_2|\mathbf{t}_2] = \left[\mathbf{U}\mathbf{W}\mathbf{V}^T \middle| +\mathbf{u}_3\right]$$
$$[\mathbf{R}_2|\mathbf{t}_2] = \left[\mathbf{U}\mathbf{W}\mathbf{V}^T \middle| -\mathbf{u}_3\right]$$
$$[\mathbf{R}_2|\mathbf{t}_2] = \left[\mathbf{U}\mathbf{W}^T\mathbf{V}^T \middle| +\mathbf{u}_3\right]$$
$$[\mathbf{R}_2|\mathbf{t}_2] = \left[\mathbf{U}\mathbf{W}^T\mathbf{V}^T \middle| -\mathbf{u}_3\right]$$

$$\mathbf{W} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

稀疏重建（Sparse Reconstruction）

- 选择四组解中三角化得到的**X**在两个相机前方数量最多的解

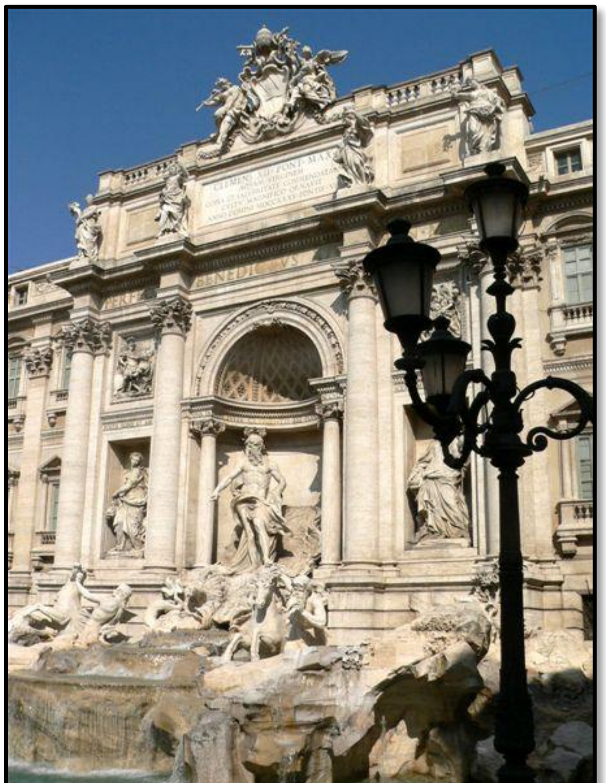


稀疏重建（Sparse Reconstruction）

- 刚才假设摄像机内参数**K**已知，实践中如何给定**K**?

稀疏重建（Sparse Reconstruction）

- 通过图片EXIF获取相机焦距、镜头型号等



```
File size      : 85111 bytes
File date      : 2005:12:16 04:17:12
Camera make    : Panasonic
Camera model   : DMC-FZ20
Date/Time      : 2005:03:19 12:52:33
Resolution     : 450 x 600
Flash used     : No
Focal length   : 6.0mm
Exposure time: 0.0012 s  (1/800)
Aperture       : f/5.6
ISO equiv.     : 80
Whitebalance : Auto
Metering Mode: matrix
Exposure       : program (auto)
```

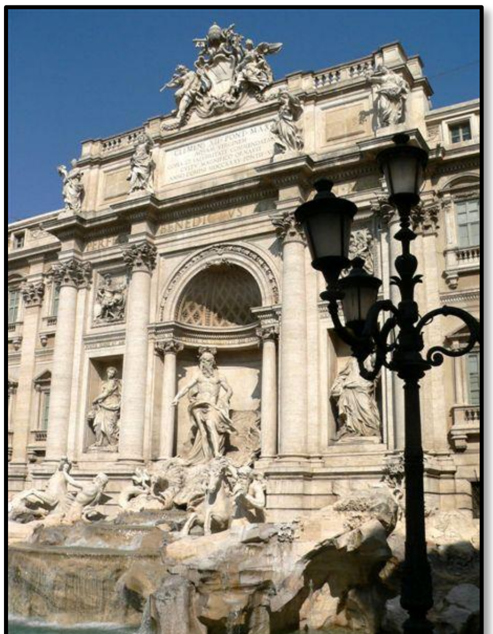
稀疏重建（Sparse Reconstruction）

- 通过镜头型号确定CCD物理尺寸

	Panasonic Lumix DMC-FZ20
Image	
Format	Compact, SLR-like
Price (street)	
Also known as	
Release Status	Discontinued
Max resolution	2560 x 1920
Low resolution	2048 x 1536, 1920 x 1080, 1600 x 1200, 1280 x 960, 640 x 480
Image ratio w:h	4:3, 16:9
Effective pixels	5.0 million
Sensor photo detectors	5.3 million
Sensor size	1/2.5 " (5.75 x 4.31 mm, 0.24 cm²)
Pixel density	20 MP/cm²
Sensor type	CCD

稀疏重建（Sparse Reconstruction）

- 刚才假设摄像机内参数**K**已知，实践中如何给定**K**?



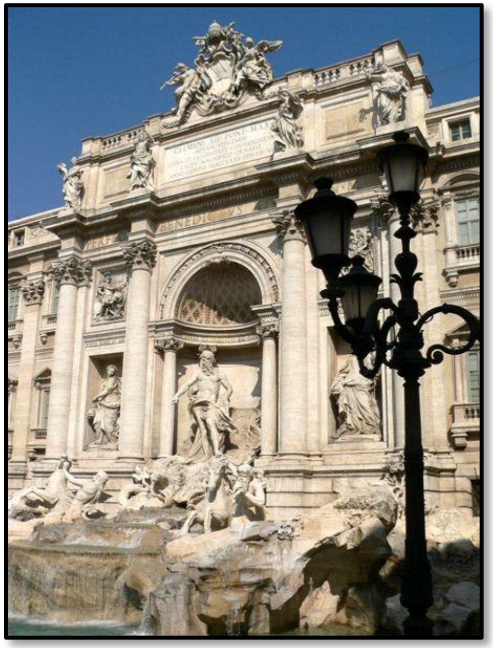
File size	85111 bytes
File date	2005:12:16 04:17:12
Camera make	**Panasonic**
Camera model	**DMC-FZ20**
Date/Time	2005:03:19 12:52:33
Resolution	**450 x 600**
Flash used	No
Focal length	**6.0mm**
Exposure time	0.0012 s (1/800)
Aperture	f/5.6
ISO equiv.	80
Whitebalance	Auto
Metering Mode	matrix
Exposure	program (auto)
Sensor size	**5.75mm**

Focal length (pixels) = Focal length (mm) x Image width (pixels) / Sensor size (mm)

= 6.0 mm x 600 pixels / 5.75 mm = 626.1 pixels

稀疏重建（Sparse Reconstruction）

- 刚才假设摄像机内参数**K**已知，实践中如何给定**K**?



File size	:	85111 bytes
File date	:	2005:12:16 04:17:12
Camera make	:	**Panasonic**
Camera model	:	**DMC-FZ20**
Date/Time	:	2005:03:19 12:52:33
Resolution	:	**450 x 600**
Flash used	:	No
Focal length	:	**6.0mm**
Exposure time	:	0.0012 s (1/800)
Aperture	:	f/5.6
ISO equiv.	:	80
Whitebalance	:	Auto
Metering Mode	:	matrix
Exposure	:	program (auto)
Sensor size	:	**5.75mm**

$$\mathbf{K}=\begin{bmatrix}626.1 & 0 & 300\\ 0 & 626.1 & 225\\ 0 & 0 & 1\end{bmatrix}$$

稀疏重建（Sparse Reconstruction）

- 以给定的**K**、通过**E**分解求解的**R**、**t**、以及三角化后的**X**为初始值，迭代求解重投影误差最小化问题：

$$\min \sum_i \left\| \mathbf{P}\mathbf{X}_i - \mathbf{x}_i \right\|^2$$

得到几何误差最小化意义下的**K**、**R**、**t**、**X**

稀疏重建（Sparse Reconstruction）

- 对比三个重投影误差最小化问题：

$$\min \sum_i \left\| \mathbf{P}\mathbf{X}_i - \mathbf{x}_i \right\|^2$$

三角化： $\min \sum_i \left\| \mathbf{P}_i \mathbf{X} - \mathbf{x}_i \right\|^2$ **x**、**K**、**R**、**t**已知，求**X**

相机标定： $\min \sum_i \left\| \mathbf{P}\mathbf{X}_i - \mathbf{x}_i \right\|^2$ **x**、**X**已知，求**K**、**R**、**t**

稀疏重建： $\min \sum_i \left\| \mathbf{P}_i \mathbf{X}_i - \mathbf{x}_i \right\|^2$ **x**已知，求**K**、**R**、**t**、**X**

- 通过线性方法求解初始值（代数误差最小化）
- 通过非线性优化迭代求精（几何误差最小化）

稀疏重建（Sparse Reconstruction）

- 重投影误差最小化问题的求解

$$\min \sum_i \left\| \mathbf{P}_i \mathbf{X}_i - \mathbf{x}_i \right\|^2$$

$$\Downarrow$$

$$\min \sum_i \sum_j \left\| \frac{\mathrm{P}_{1:2}^i \mathbf{X}_j}{\mathrm{P}_3^i \mathbf{X}_j} - \begin{bmatrix} x_{ij} \\ y_{ij} \end{bmatrix} \right\|^2$$

$$\Downarrow$$

$$\min \sum_i \sum_j \left\| f_i(\hat{\mathbf{X}}) - b_{ij} \right\|^2$$

稀疏重建（Sparse Reconstruction）

- 重投影误差最小化问题的求解

$$\min \sum_i \sum_j \left\| f_i(\hat{\mathbf{X}}) - b_{ij} \right\|^2$$

三角化：　$\hat{\mathbf{X}}$ 包括$\mathbf{X}$

相机标定：$\hat{\mathbf{X}}$ 包括$\mathbf{K}$、$\mathbf{R}$、$\mathbf{t}$

稀疏重建：$\hat{\mathbf{X}}$ 包括$\mathbf{K}$、$\mathbf{R}$、$\mathbf{t}$、$\mathbf{X}$

$$\Downarrow$$

$$\min \left\| \mathrm{F}(\hat{\mathbf{X}}) - \mathrm{b} \right\|^2 \quad \mathrm{F}(\hat{\mathbf{X}}) = \begin{bmatrix} \vdots \\ f_i(\hat{\mathbf{X}}) \\ \vdots \end{bmatrix} \quad \mathrm{b} = \begin{bmatrix} \vdots \\ \mathrm{b}_{ij} \\ \vdots \end{bmatrix}$$

稀疏重建（Sparse Reconstruction）

- 重投影误差最小化问题的求解

$$\min\left\|F(\hat{\mathbf{X}})-b\right\|^2$$

重投影误差最小化是一个非线性最小二乘问题，求解非线性最小二乘问题的方法：

- 非线性最小二乘是非线性优化的一类特殊形式；
- 针对一般非线性优化的梯度下降法（1阶）、牛顿法（2阶）；
- 针对非线性最小二乘的高斯牛顿法、LM法（Levenberg-Marquardt）。

稀疏重建（Sparse Reconstruction）

- 重投影误差最小化问题的求解

$$\min \left\| F(\hat{\mathbf{X}}) - b \right\|^2$$

非线性最小二乘问题迭代优化的基本思路：

1）给定初始值；

2）开始迭代优化

—选择最优移动方向使目标函数值下降最快；

—以一定步长沿最优方向移动当前值；

—如果两次迭代间目标函数值差异小于阈值或迭代次数超出阈值，则转步骤3），否则返回2）；

3）迭代结束，输出当前值。

稀疏重建（Sparse Reconstruction）

$$\min\left\|\mathrm{F}(\hat{\mathbf{X}})-\mathrm{b}\right\|^2$$

Gradient Descent方向为： $\delta(\hat{\mathbf{X}})=-\mathrm{J}(\hat{\mathbf{X}})^T\,\mathrm{r}(\hat{\mathbf{X}})$

Newton方向为： $\delta(\hat{\mathbf{X}})=-(\mathrm{J}(\hat{\mathbf{X}})^T\,\mathrm{J}(\hat{\mathbf{X}})+\mathrm{S}(\hat{\mathbf{X}}))^{-1}\,\mathrm{J}(\hat{\mathbf{X}})^T\,\mathrm{r}(\hat{\mathbf{X}})$

Gauss-Newton方向为： $\delta(\hat{\mathbf{X}})=-(\mathrm{J}(\hat{\mathbf{X}})^T\,\mathrm{J}(\hat{\mathbf{X}}))^{-1}\,\mathrm{J}(\hat{\mathbf{X}})^T\,\mathrm{r}(\hat{\mathbf{X}})$

阻尼高斯牛顿方向为： $\delta(\hat{\mathbf{X}})=-(\mathrm{J}(\hat{\mathbf{X}})^T\,\mathrm{J}(\hat{\mathbf{X}})+\lambda\mathbf{I})^{-1}\,\mathrm{J}(\hat{\mathbf{X}})^T\,\mathrm{r}(\hat{\mathbf{X}})$

$$\mathrm{J}(\hat{\mathbf{X}})=\frac{\partial\,\mathrm{F}(\hat{\mathbf{X}})}{\partial\hat{\mathbf{X}}}$$

Jacobian矩阵（一阶导数）

$$\mathrm{r}(\hat{\mathbf{X}})=\mathrm{F}(\hat{\mathbf{X}})-\mathrm{b}$$

残差向量

$$\mathrm{H}(\hat{\mathbf{X}})=\mathrm{J}(\hat{\mathbf{X}})^T\,\mathrm{J}(\hat{\mathbf{X}})+\mathrm{S}(\hat{\mathbf{X}})$$

Hessian矩阵（二阶导数）

稀疏重建（Sparse Reconstruction）

- 阻尼高斯牛顿法（Damped Gauss-Newton Method）

高斯牛顿法的特点：

- 高斯牛顿法的收敛速度比梯度下降法快的多（超线性收敛，近似二阶收敛）；
- 高斯牛顿法不能保证收敛（梯度下降法和牛顿法收敛）。

阻尼高斯牛顿法：$\delta(\hat{\mathbf{X}},\lambda) = -(\mathrm{J}(\hat{\mathbf{X}})^T \mathrm{J}(\hat{\mathbf{X}}) + \lambda \mathbf{I})^{-1} \mathrm{J}(\hat{\mathbf{X}})^T \mathrm{r}(\hat{\mathbf{X}})$

$\lim_{\lambda\to 0} \delta(\hat{\mathbf{X}},\lambda)$ 则成为高斯牛顿法

$\lim_{\lambda\to \infty} \delta(\hat{\mathbf{X}},\lambda)$ 则成为梯度下降法

稀疏重建（Sparse Reconstruction）

- Levenberg-Marquardt Method (LM法)

LM法通过启发式方法在每一步动态调整λ:

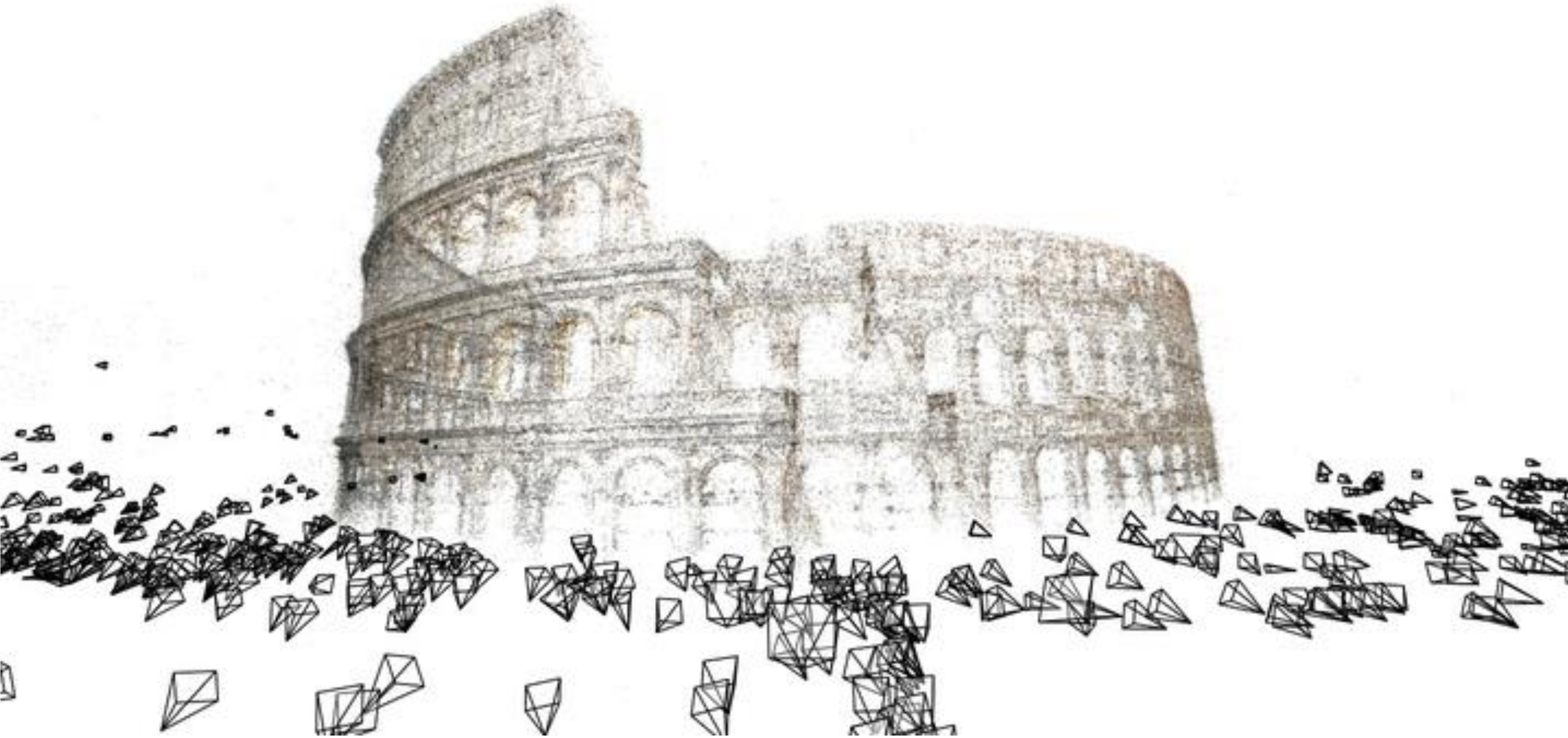
$$\delta(\hat{\mathbf{X}}, \lambda) = -(\mathrm{J}(\hat{\mathbf{X}})^T \mathrm{J}(\hat{\mathbf{X}}) + \lambda \mathbf{I})^{-1} \mathrm{J}(\hat{\mathbf{X}})^T \mathrm{r}(\hat{\mathbf{X}})$$

如果误差减少，则令$\lambda \leftarrow 0.1\lambda$

如果误差增大，则令$\lambda \leftarrow 10\lambda$

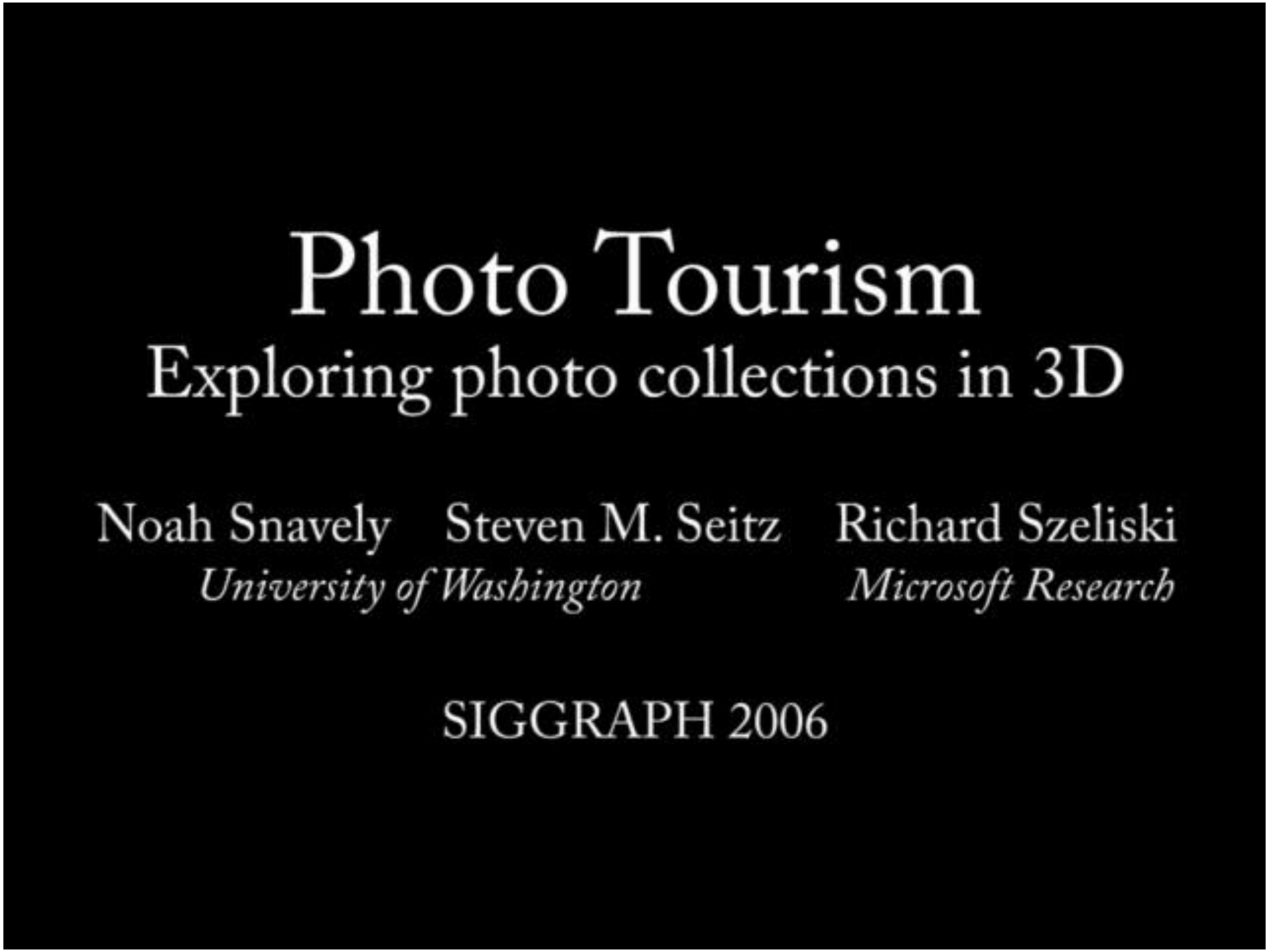
- LM法是一种启发式的阻尼高斯牛顿法，在计算机视觉中广泛使用。
- 使用LM求解重投影误差最小化的方法称为Bundle Adjustment (捆绑调整，摄影测量中称为光束平差)

实践：Photo Tourism



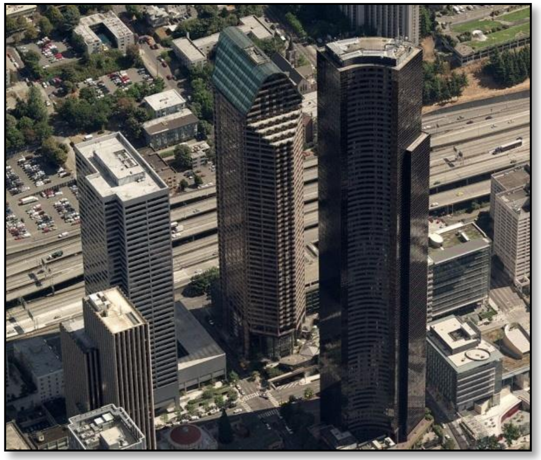
- N. Snavely, S. Seitz, R. Szeliski, *Photo tourism: Exploring photo collections in 3D*, Siggraph 2006.
- S. Agarwal, N. Snavely, I. Simon, S. Seitz, R. Szeliski, *Building Rome in a Day*, ICCV 2009.

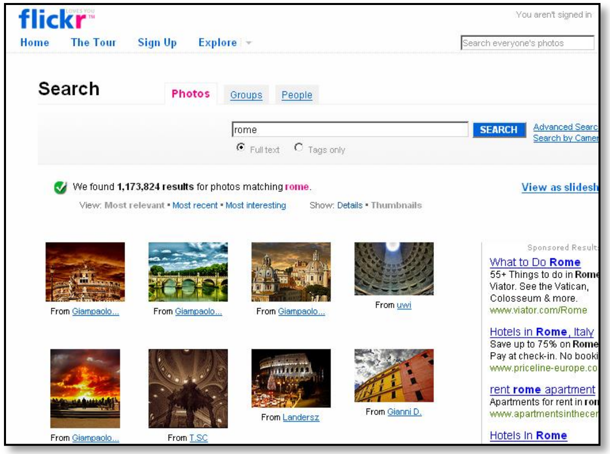
实践：Photo Tourism



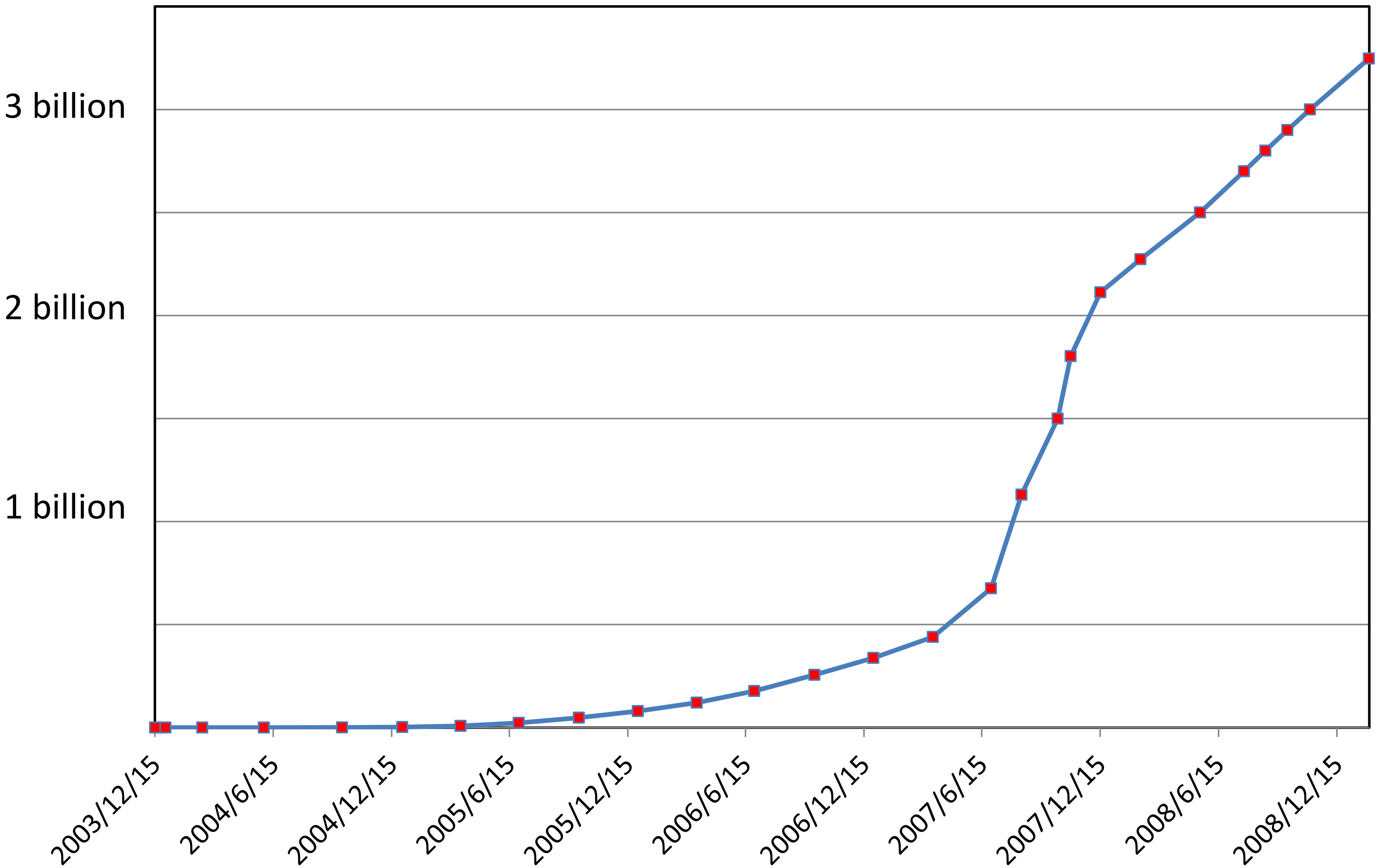
The world in photos

- There are ***billions*** of photos online
- Photographic record of the surface of the earth
- Photo sharing on a massive scale





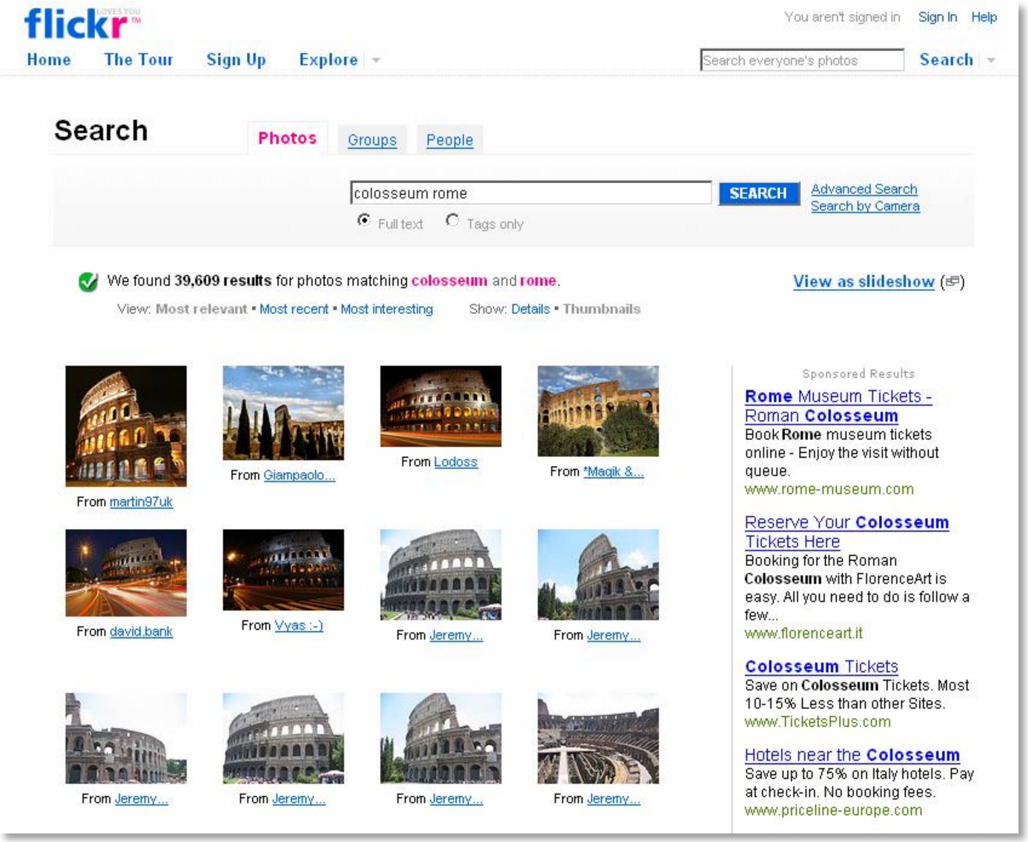
Flickr



> **3.6 billion** photos on Flickr, > **7.2 billion** on Photobucket, > **15 billion** on Facebook

Internet structure from motion

- Input: collection of photos resulting from Internet search



Internet structure from motion

- Input: collection of photos resulting from Internet search

- Input characteristics:
  - Taken by many different people and cameras

Motorola RAZR

Nikon D3

Internet structure from motion

- Input: collection of photos resulting from Internet search

- Input characteristics:
  - Taken at many different times of day, year, century

Internet structure from motion

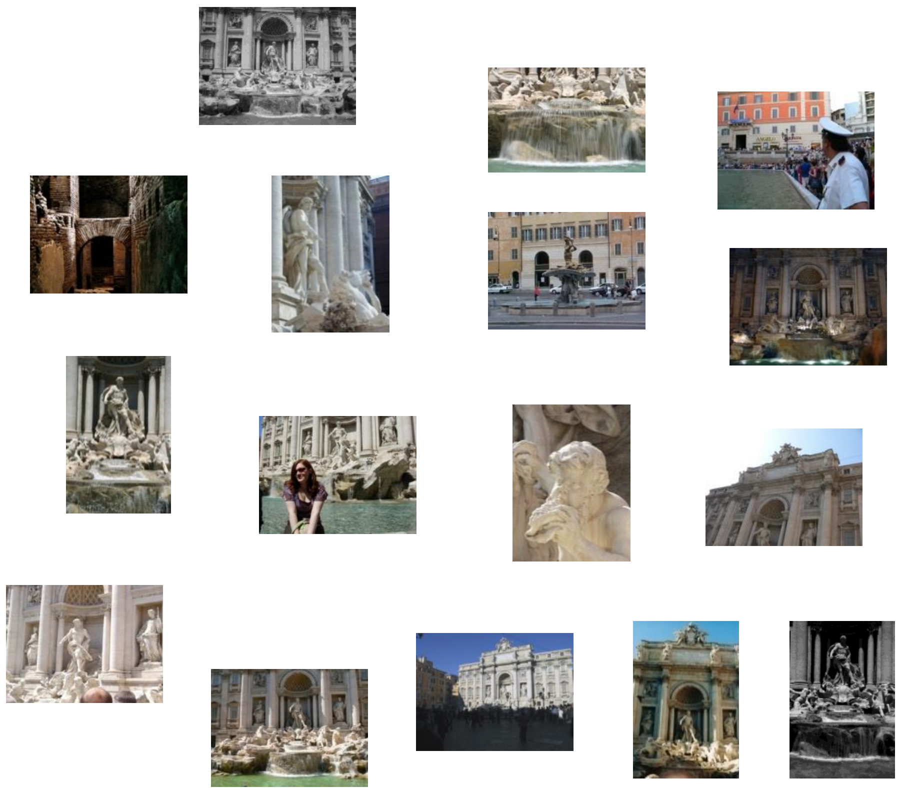
- Input: collection of photos resulting from Internet search
- Input characteristics:
  - Given in essentially random order

Two important breakthroughs

- Advances in wide-baseline feature matching (e.g., SIFT)
- Advances in multi-view geometry techniques

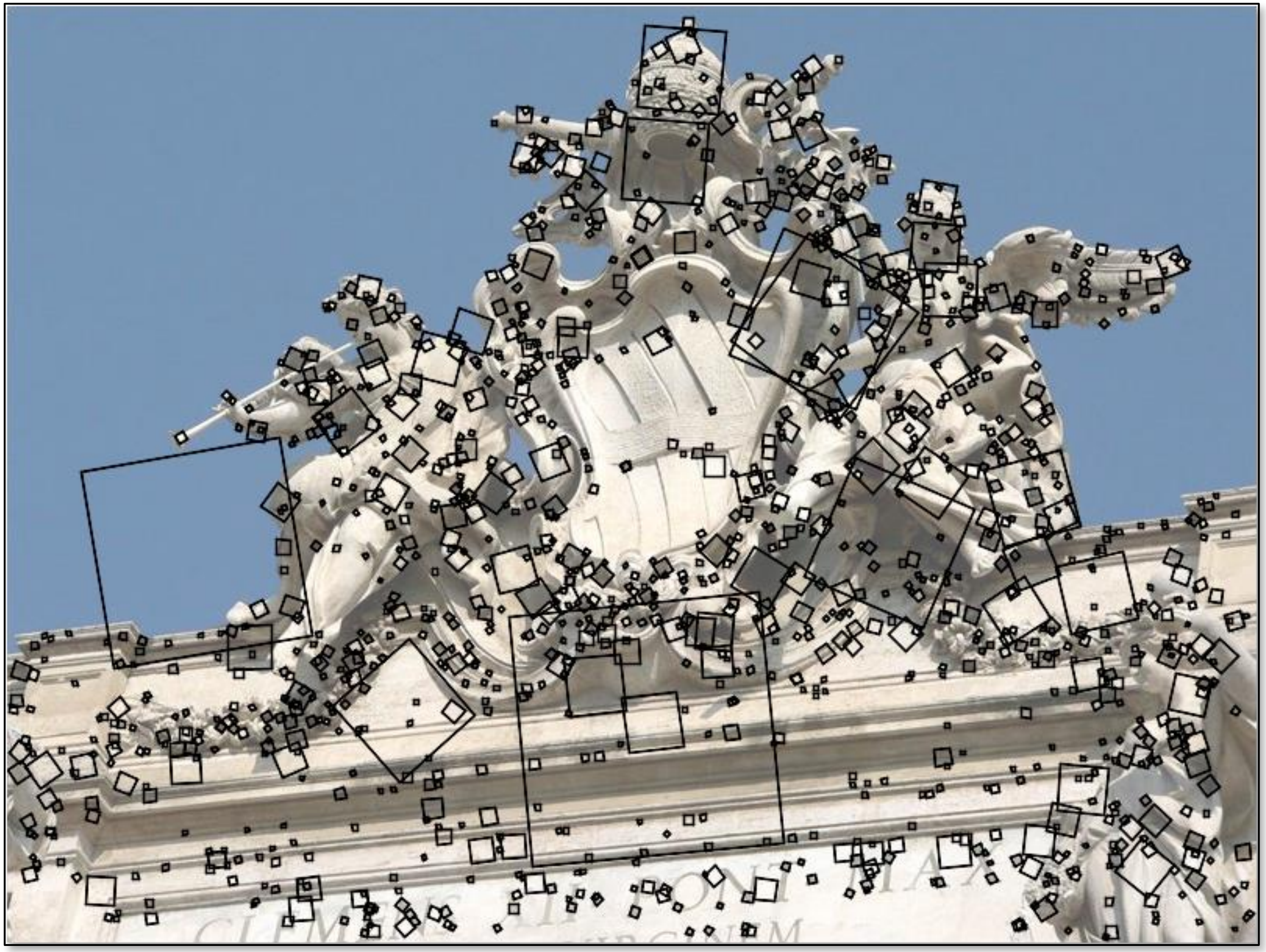
Feature detection

Detect features using SIFT [Lowe, IJCV 2004]



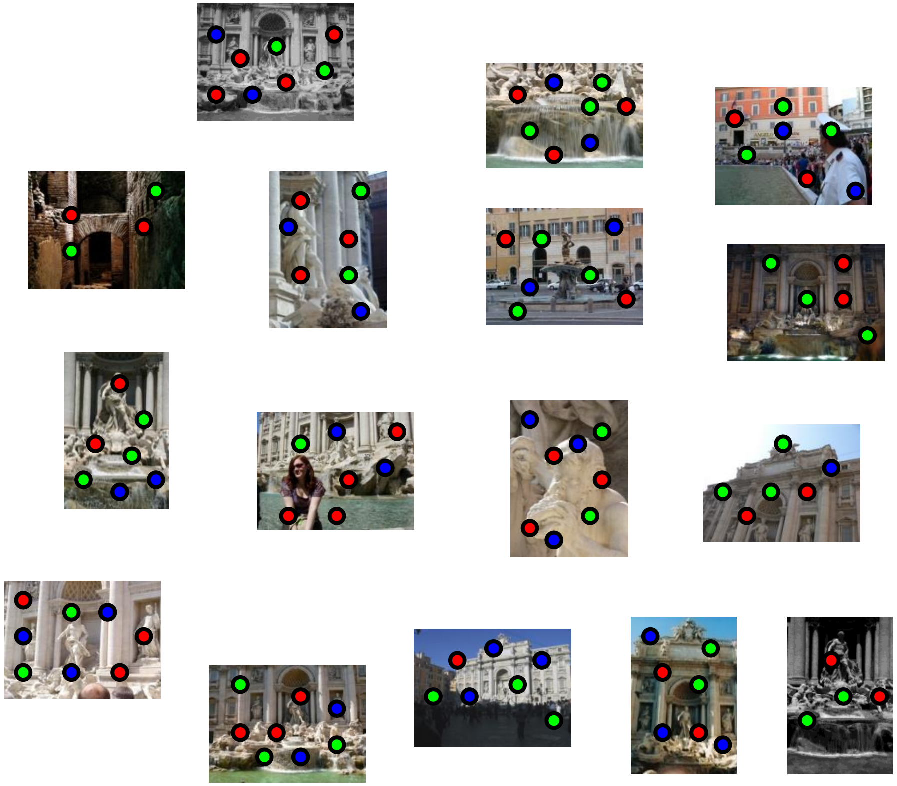
Feature detection

Detect features using SIFT [Lowe, IJCV 2004]



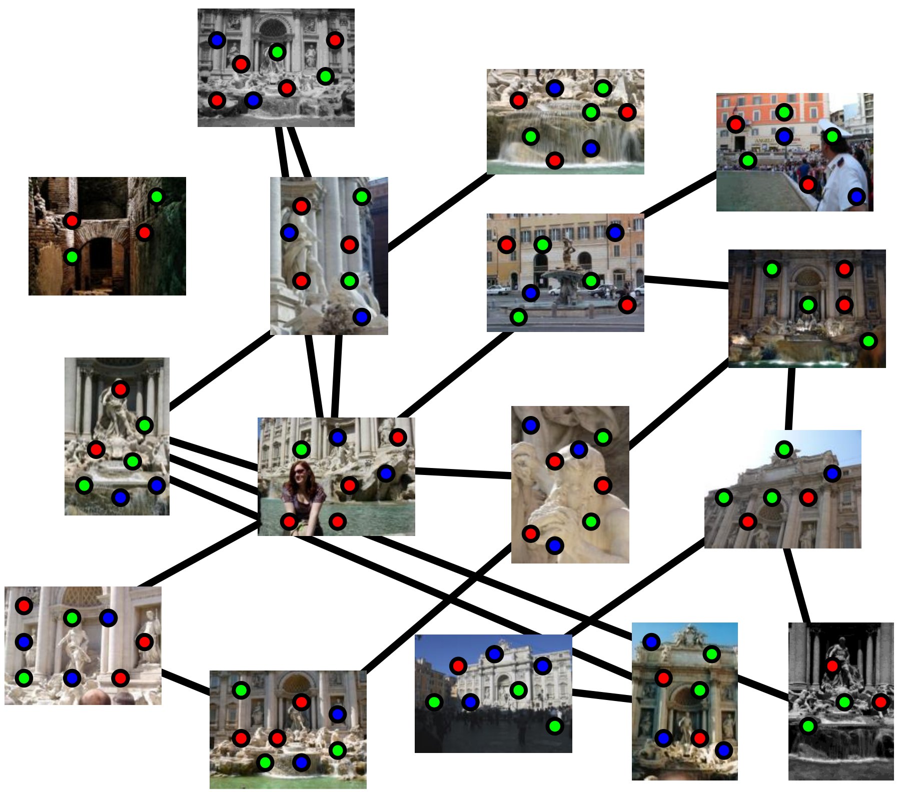
Feature detection

Detect features using SIFT [Lowe, IJCV 2004]



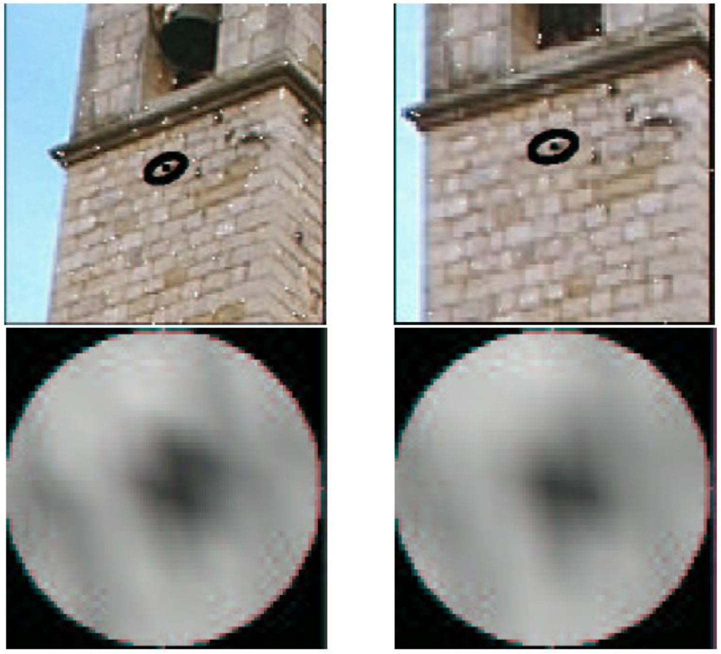
Wide-baseline feature matching

Match features between each pair of images



Wide-baseline feature matching

- Standard approach for pairwise matching:
  - For each feature in image A:
  - Find the feature with the closest descriptor in image B



From Schaffalitzky and Zisserman ’02

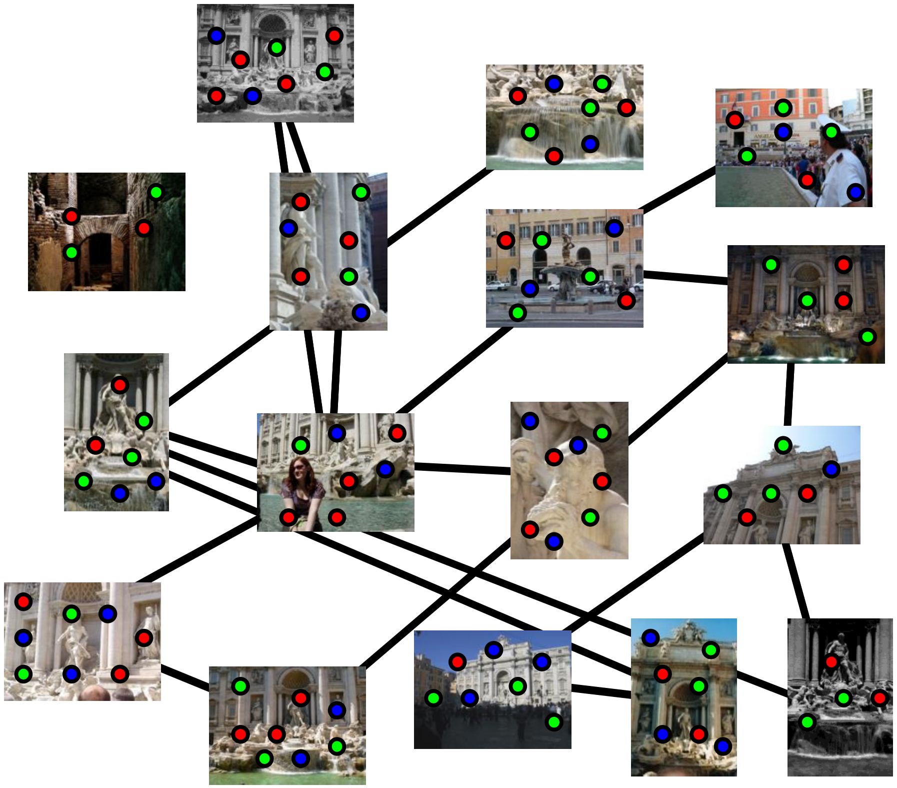
Wide-baseline feature matching

- Compare the distance to the closest feature to the distance to the second closest feature

- If the ratio of distances is less than a threshold, keep the feature
- Why the ratio test?
  - Eliminates hard-to-match repeated features
  - Distances in SIFT space seem to be non-uniform

Wide-baseline feature matching

Refine matching using RANSAC + 8-point algorithm to estimate fundamental matrices between pairs



The RANSAC Song (http://danielwedge.com/ransac/)

The power of SIFT

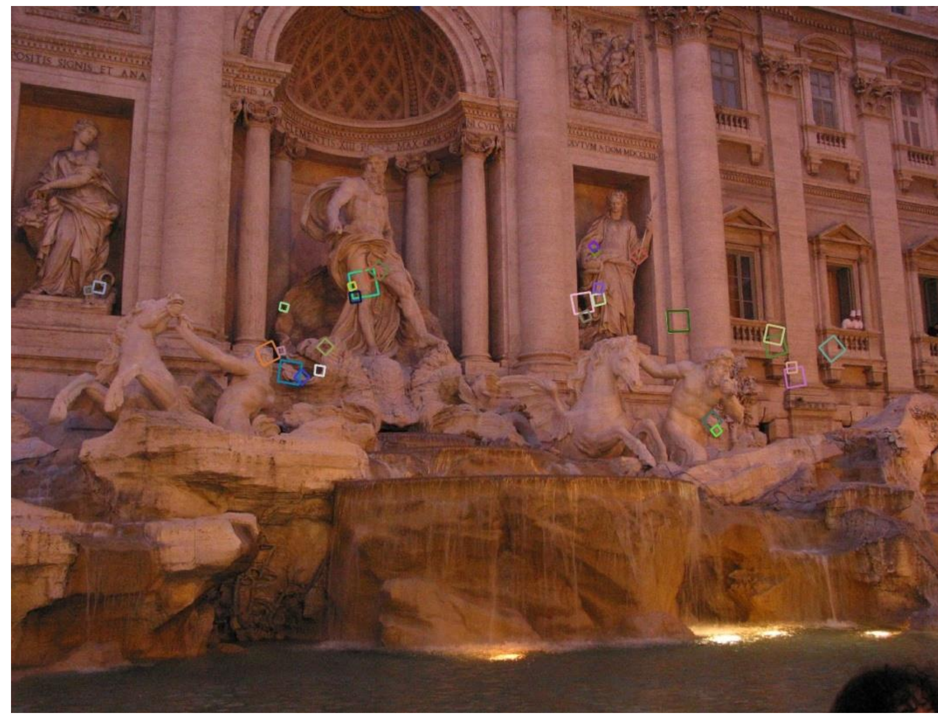
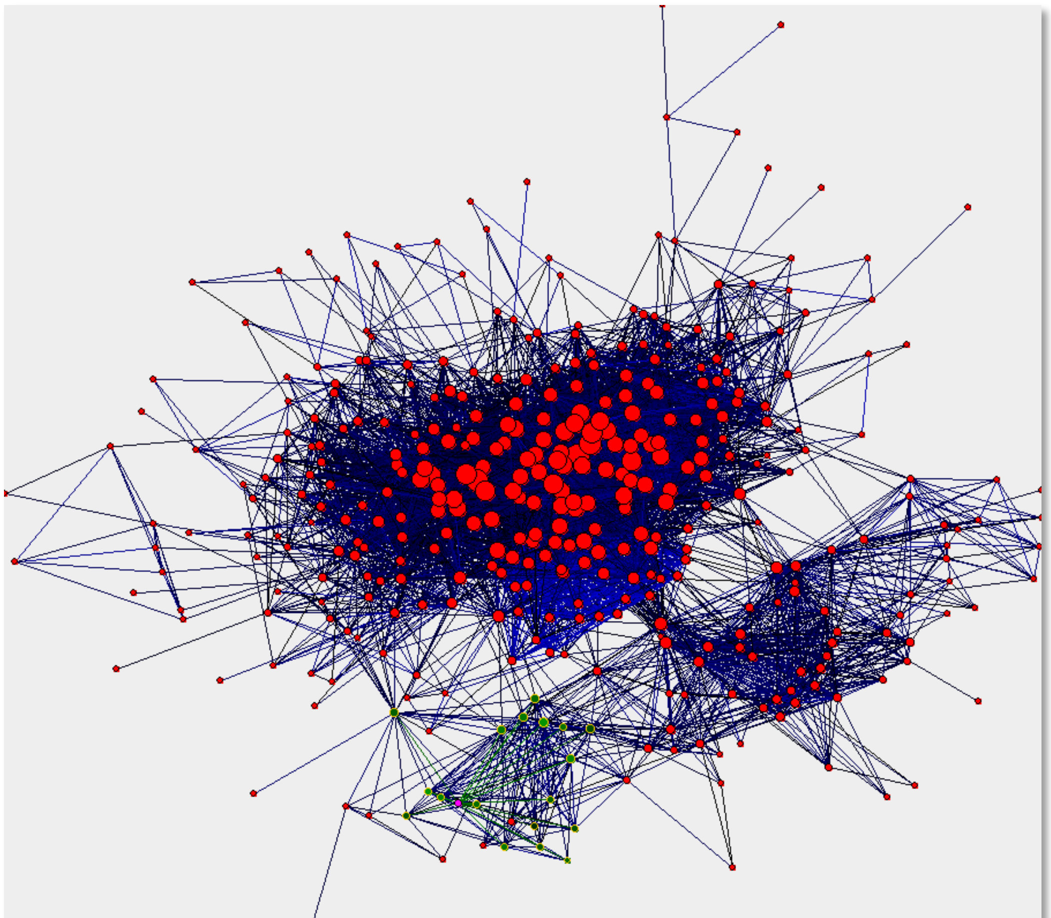


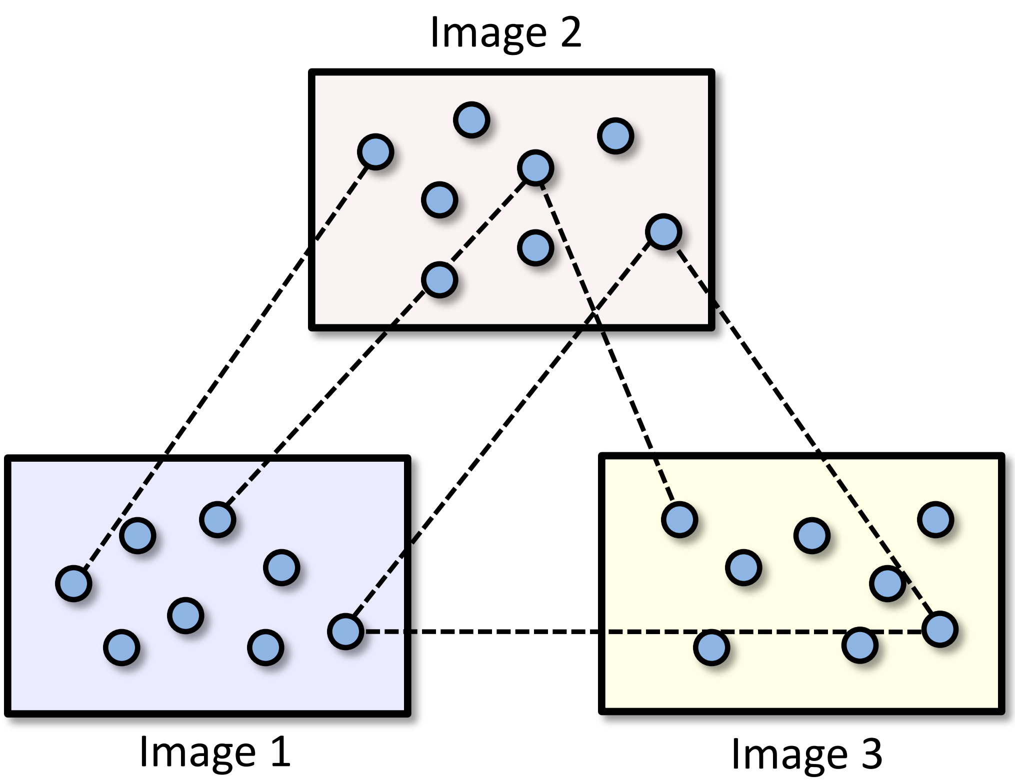
Image connectivity graph



(graph layout produced using the Graphviz toolkit: http://www.graphviz.org/)

From pairwise matches to tracks

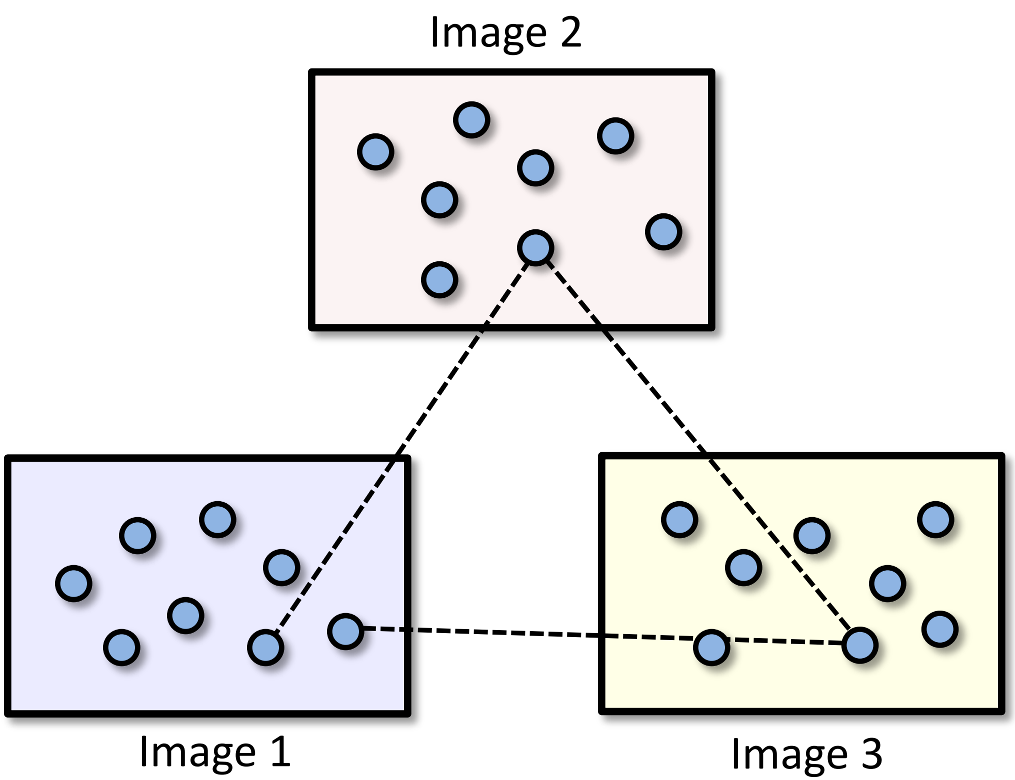
- Once we have pairwise matches, next step is to link up matches to form *tracks*



- Each track is a connected component of the pairwise feature match graph
- Each track will eventually grow up to become a 3D point

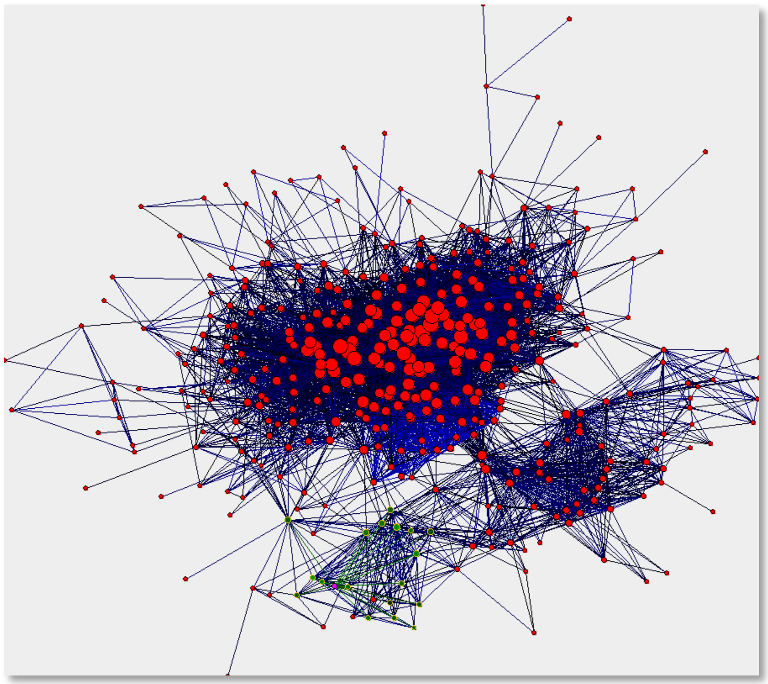
From pairwise matches to tracks

- Once we have pairwise matches, next step is to link up matches to form *tracks*



- Some tracks might be *inconsistent*
- We remove the features from the troublesome images

Image connectivity post track generation



Raw image matches

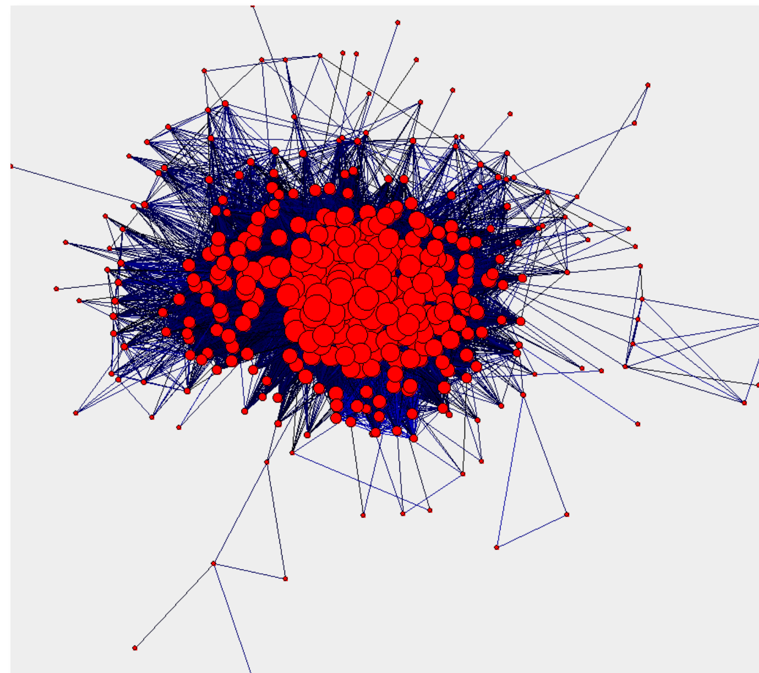


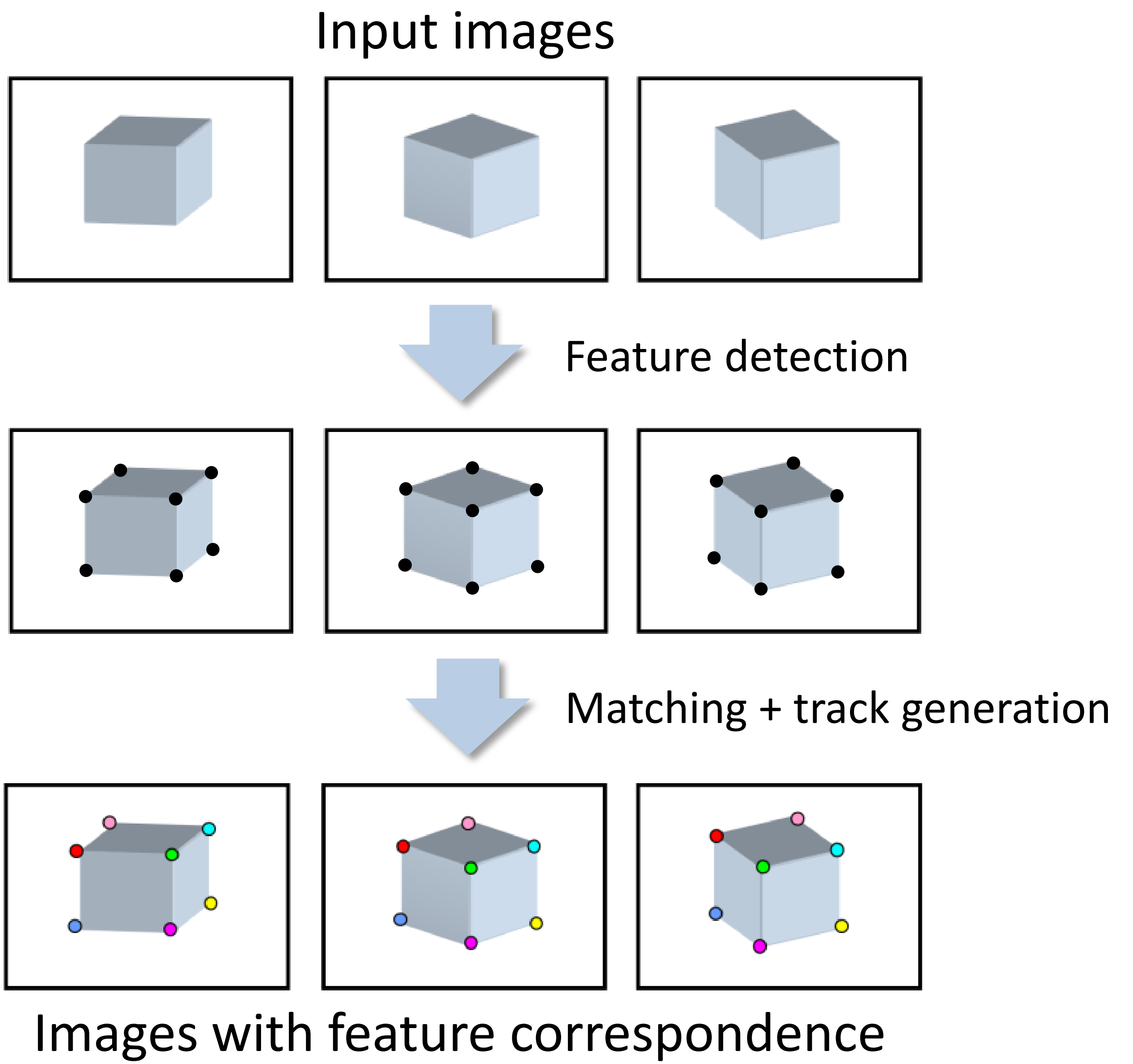
Image matches after track generation

The power of transitivity

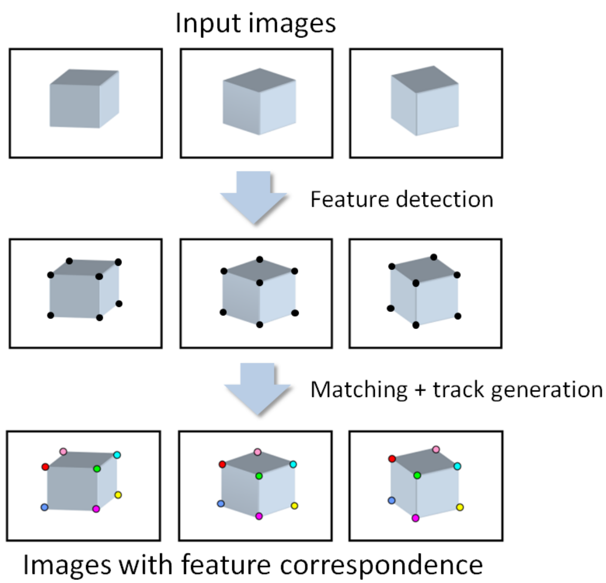
… but most tracks are short

- Example image collection with 3,000 images:
  - 1,546,612 total tracks
  - 79% have length 2
  - 90% have length <= 3
  - 98% have length <= 10
  - Longest track: 385 features

The story so far...



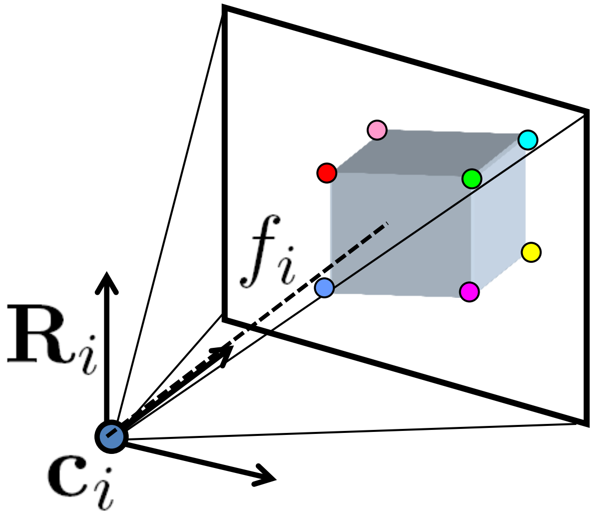
The story so far...



- Next step:
  - Use structure from motion to solve for geometry (cameras and points)
- First: what are cameras and points?

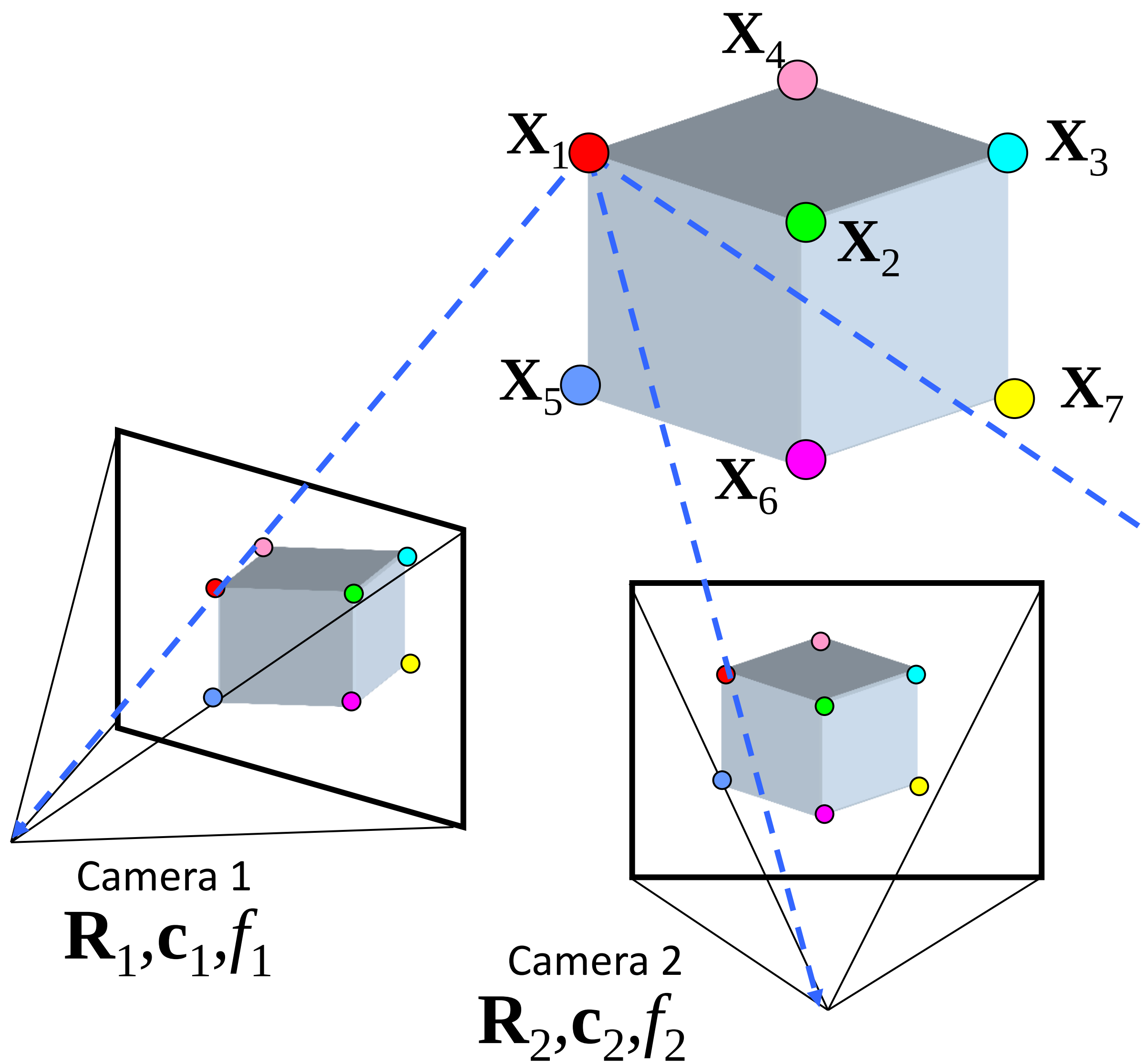
Points and cameras

- Point: 3D position in space ($\mathbf{X}_j$)
- Camera (C_i):
  - A 3D position ($\mathbf{c}_i$)
  - A 3D orientation ($\mathbf{R}_i$)
  - Intrinsic parameters (**focal length**, aspect ratio, …)
  - 7 parameters (3+3+1) in total



$$\mathbf{c}_i = -\mathbf{R}_i^T \mathbf{t}_i$$

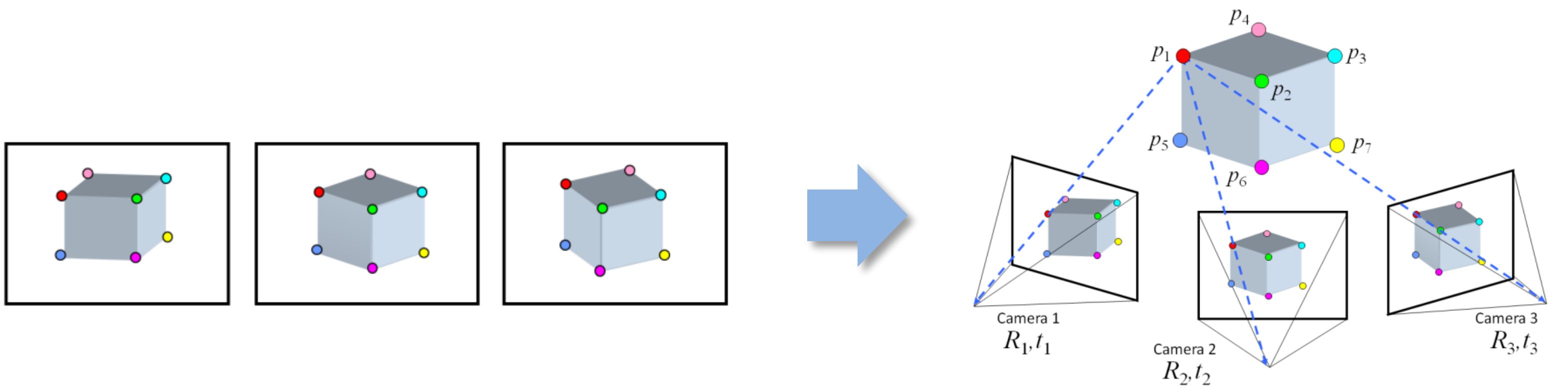
Structure from motion



minimize

$$g\,(\mathbf{R}, \mathbf{C}, \mathbf{F}, \mathbf{X})$$

Solving structure from motion

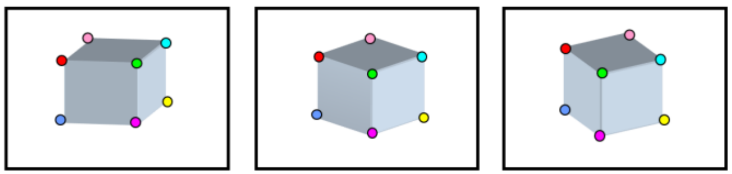


Inputs: feature tracks

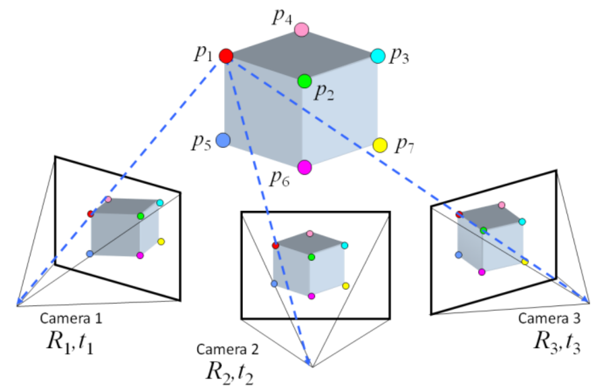
Outputs: 3D cameras and points

- How do we solve the SfM problem?
- Challenges:
  – Large number of parameters (1000's of cameras, millions of points)
  – Very non-linear objective function

Solving structure from motion



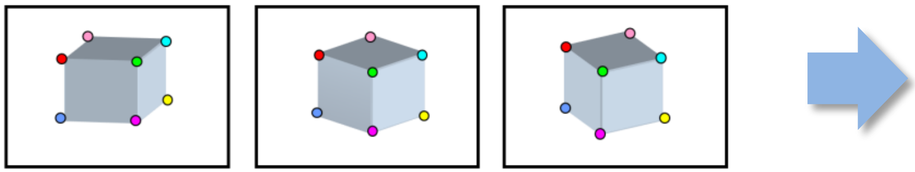
Inputs: feature tracks

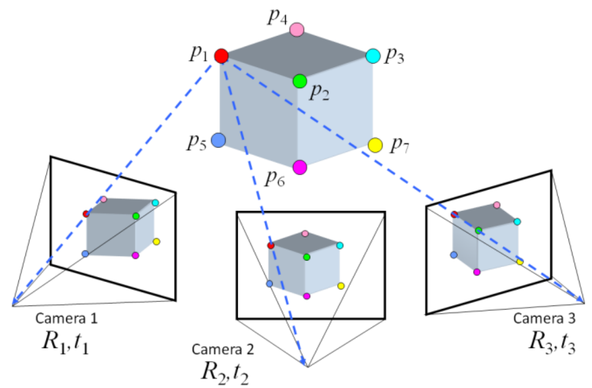


Outputs: 3D cameras and points

- Important tool: Bundle Adjustment [Triggs *et al.* '00]
  - Joint non-linear optimization of both cameras and points
  - Very powerful, elegant tool
- The bad news:
  - Starting from a random initialization is very likely to give the wrong answer
  - Difficult to initialize all the cameras at once

Solving structure from motion



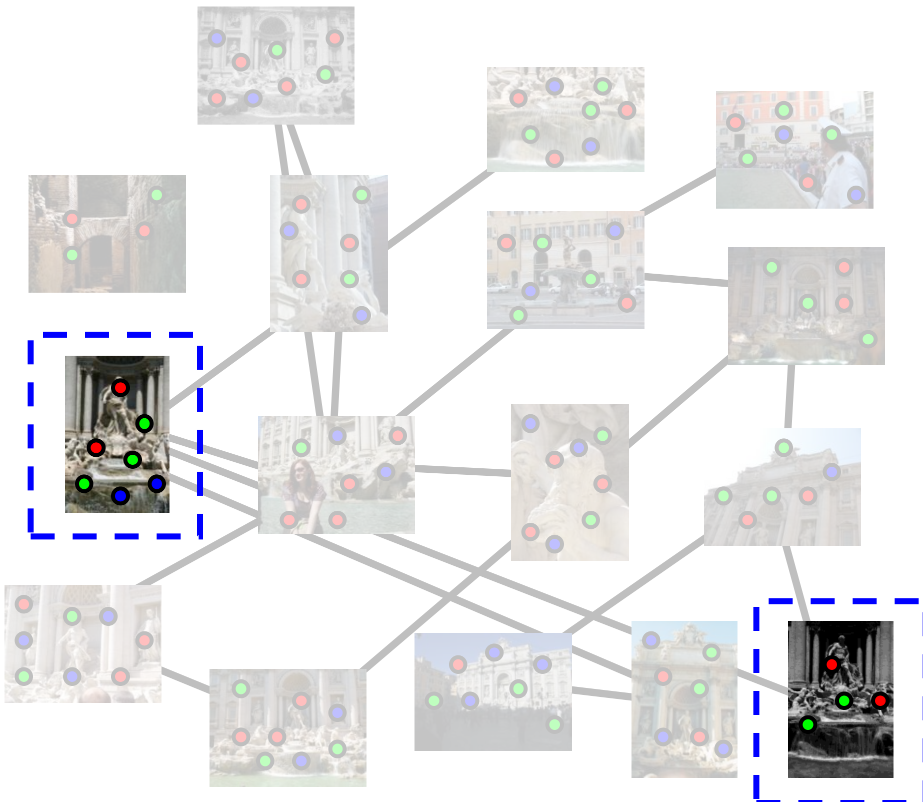


Inputs: feature tracks

Outputs: 3D cameras and points

- The good news:
  - Structure from motion with two cameras is (relatively) easy
  - Once we have an initial model, it's easy to add new cameras
- Idea:
  - Start with a small seed reconstruction, and grow

Incremental SfM



- Automatically select an initial pair of images

Incremental SfM

Incremental SfM

Incremental SfM: Algorithm

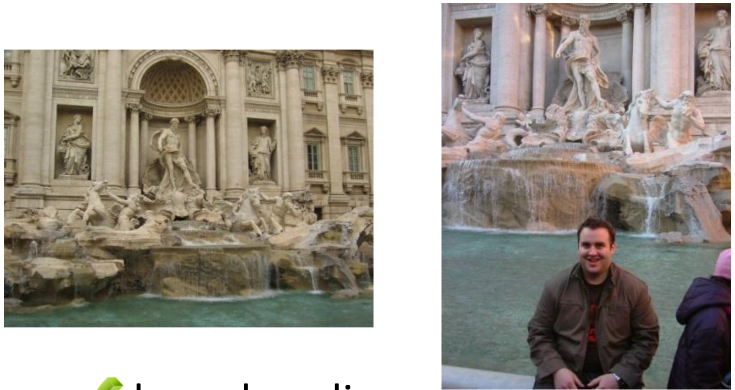
1. Pick a strong initial pair of images
2. Initialize the model using two-frame SfM
3. While there are connected images remaining:
   a. Pick the image which sees the most existing 3D points
   b. Estimate the pose of that camera
   c. Triangulate any new points
   d. Run bundle adjustment

1. Picking the initial pair

- We want a pair with many matches, but which has as large a baseline as possible

✓ lots of matches
✗ small baseline

✓ large baseline
✗ very few matches



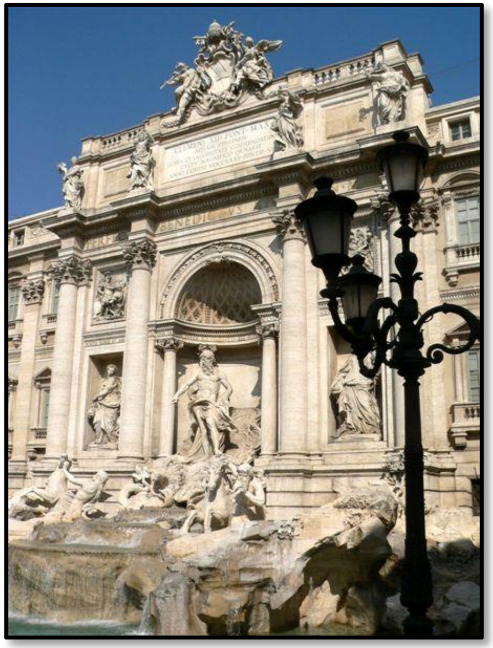
✓ large baseline
✓ lots of matches

2. Two-frame reconstruction

- Input: two images with correspondence
- Output: camera parameters, 3D points

- In general, there can be ambiguities if the cameras are uncalibrated (camera intrinsics are unknown)
- We assume that the only intrinsic parameter is an unknown focal length

Finding calibration information

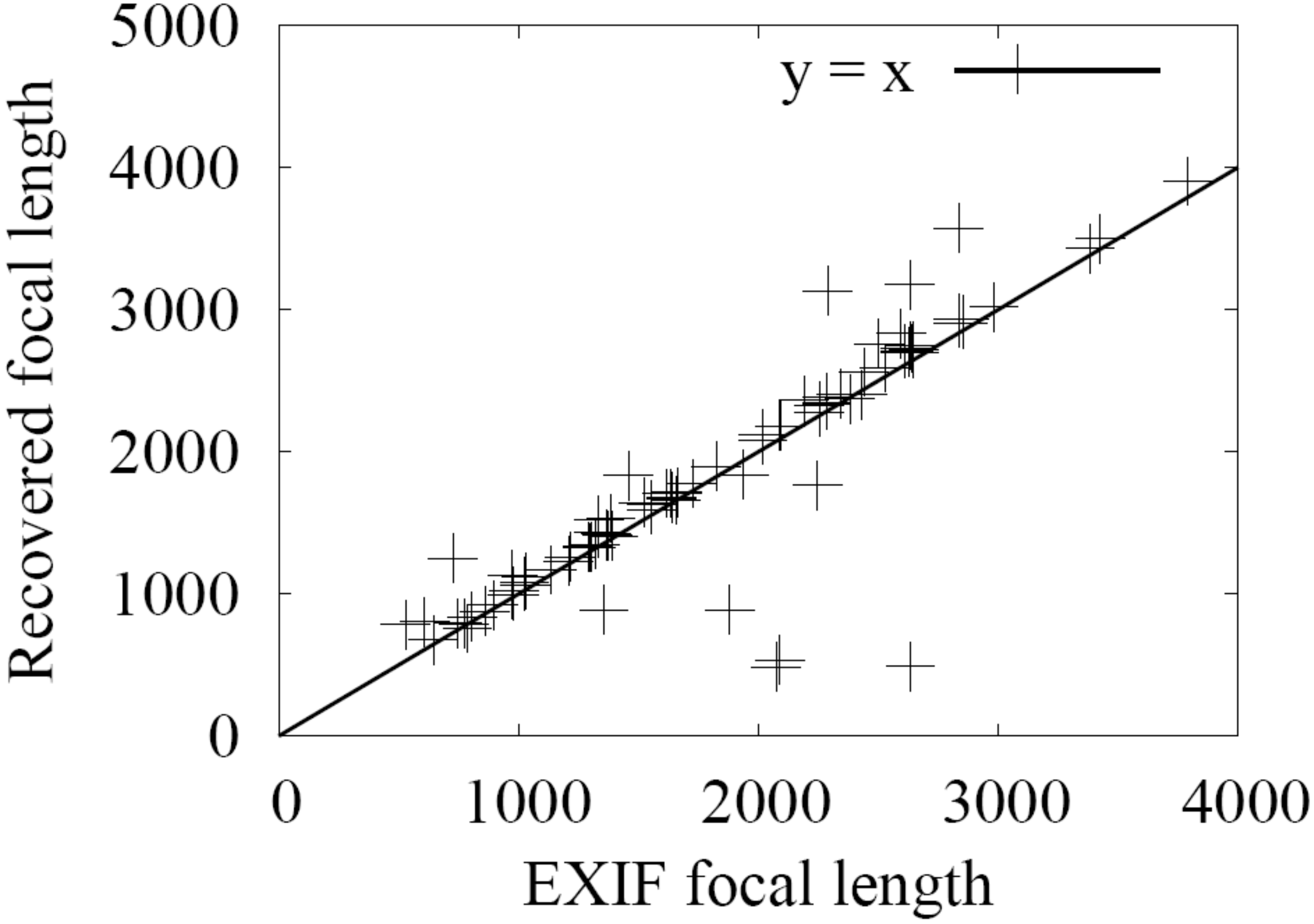


```
File size     : 85111 bytes
File date     : 2005:12:16 04:17:12
Camera make   : Panasonic
Camera model  : DMC-FZ20
Date/Time     : 2005:03:19 12:52:33
Resolution    : 450 x 600
Flash used    : No
Focal length  : 6.0mm
Exposure time: 0.0012 s  (1/800)
Aperture      : f/5.6
ISO equiv.    : 80
Whitebalance  : Auto
Metering Mode: matrix
Exposure      : program (auto)
Sensor size   : 5.75mm
```

Focal length (pixels) = Focal length (mm) x Image width (pixels) / Sensor size (mm)

= 6.0 mm x 600 pixels / 5.75 mm = 626.1 pixels

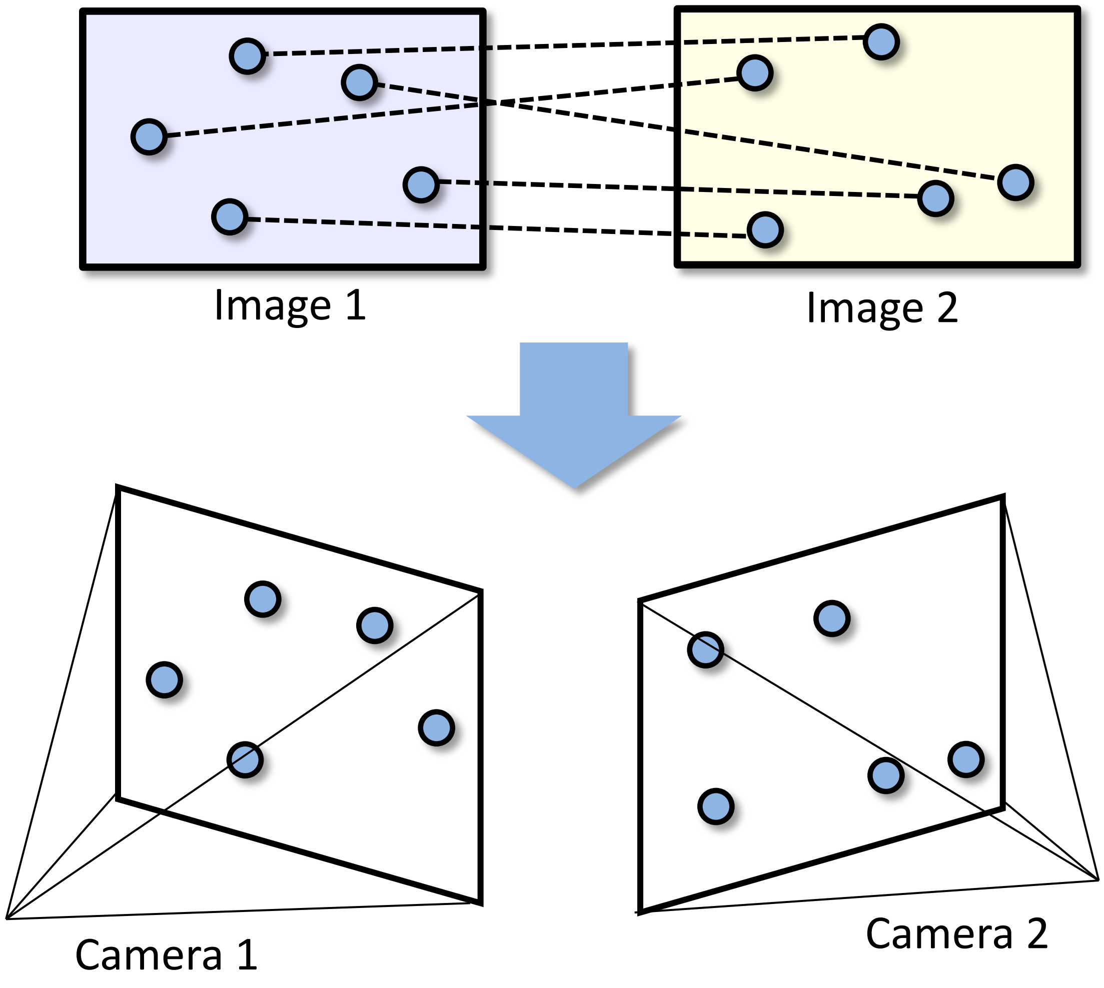
How good are Exif tags?



2. Two-view reconstruction

- Two-view SfM: Given two calibrated images with corresponding points, compute the camera and point positions
- Solved by finding the essential matrix between the images
- Best approach is the 5-point algorithm (as opposed to the 6-, 7-, or 8- point algorithms)

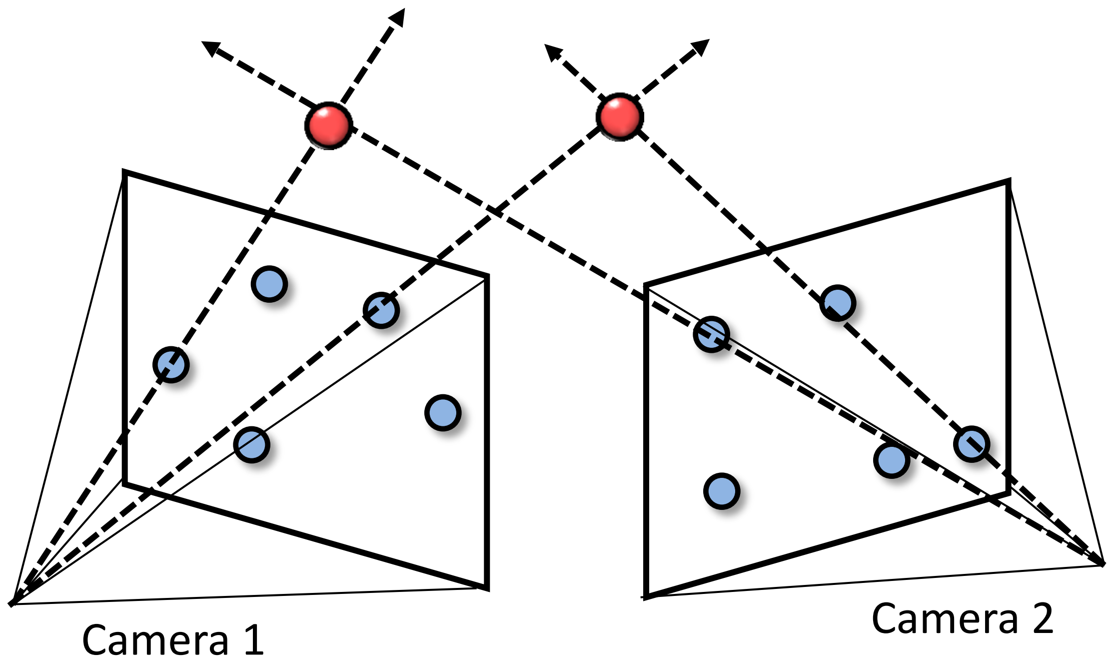
Five-point algorithm



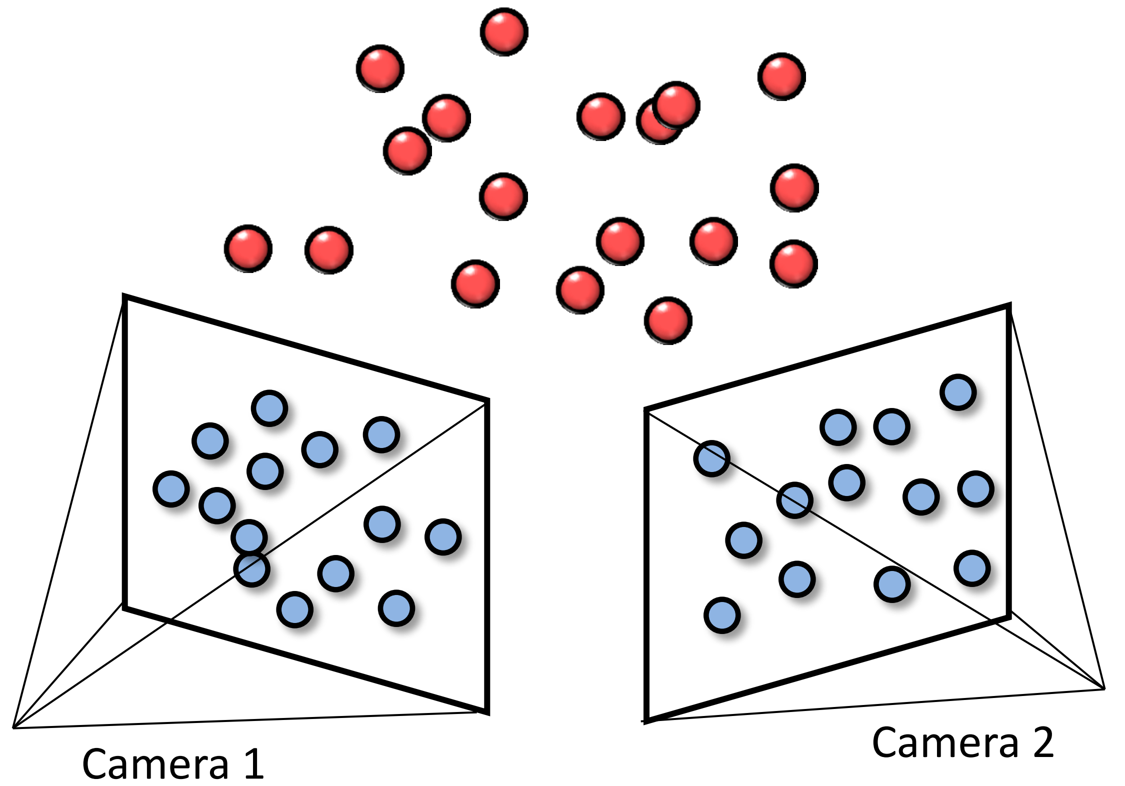
Five-point algorithm

- First practical solution to the 5-point algorithm: [Nister, “An efficient solution to the 5-point relative pose problem,” PAMI ’04]
- See also:
  - [Li and Hartley, “Five-Point Motion Estimation Made Easy,” ICPR ’06]

Two-view reconstruction

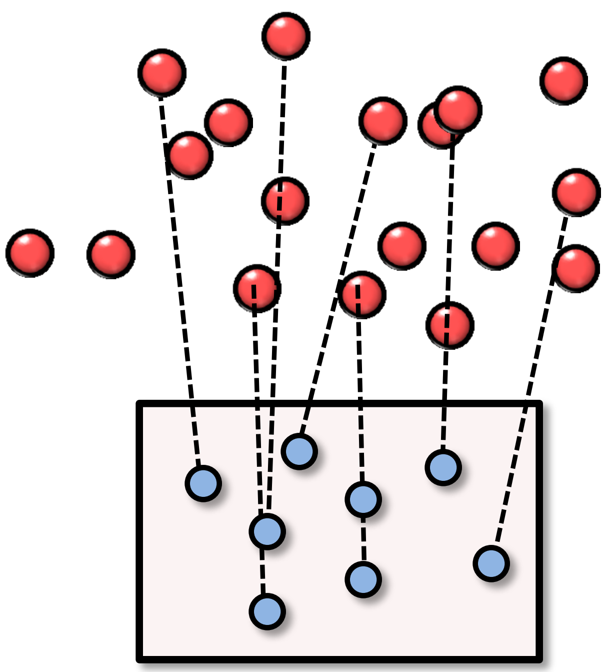


Two-view reconstruction

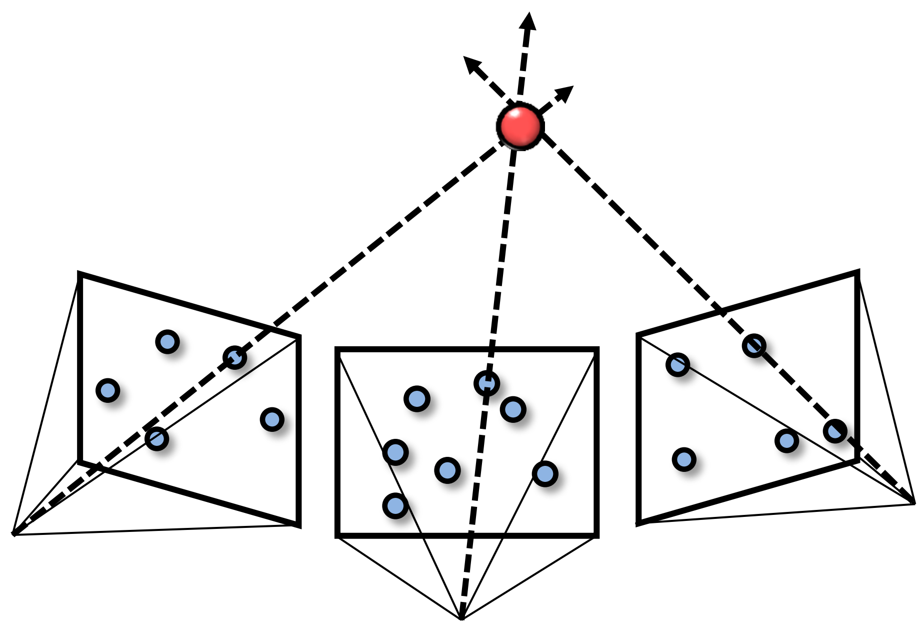


3bc. Pose estimation and Triangulation

- Next step: grow the reconstruction by adding another image, triangulating new points



Pose estimation: 2D -> 3D



n-view triangulation

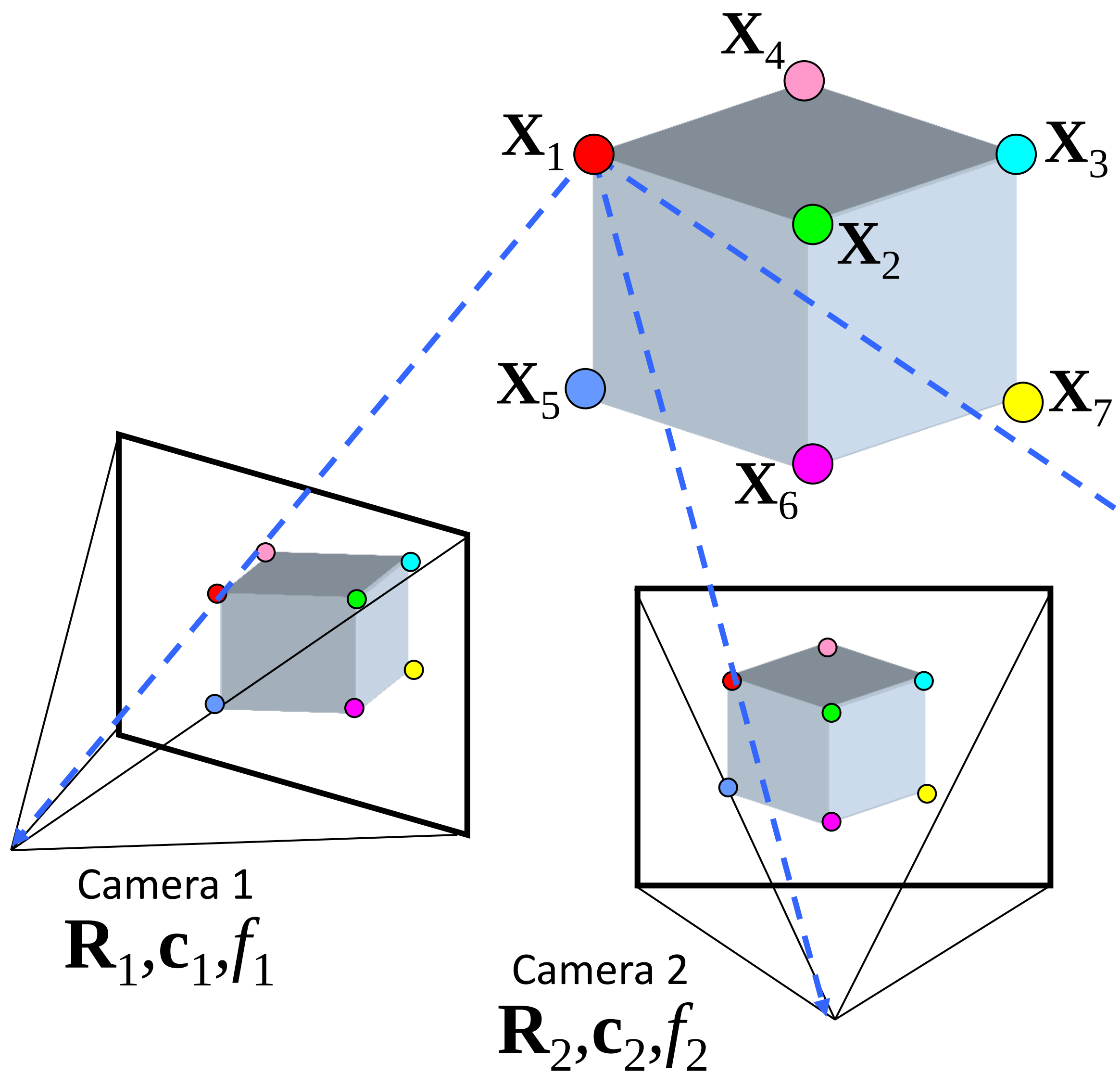
3bc. Pose estimation and triangulation

- Next step: grow the reconstruction by adding another image, triangulating new points
- Both of these problems can be solved approximately using linear systems (Direct Linear Transformation (DLT))

3bc. Pose estimation and Triangulation

- In practice, multiple images can be added at once
- If the highest-matching image has N matches, add all images with at least 0.75 N matches (or at least 500 matches)

3d. Bundle adjustment



minimize

$$g(\mathbf{R},\mathbf{C},\mathbf{F},\mathbf{X})$$

3d. Bundle adjustment

- Given:
  - Vectors of cameras and 3D points

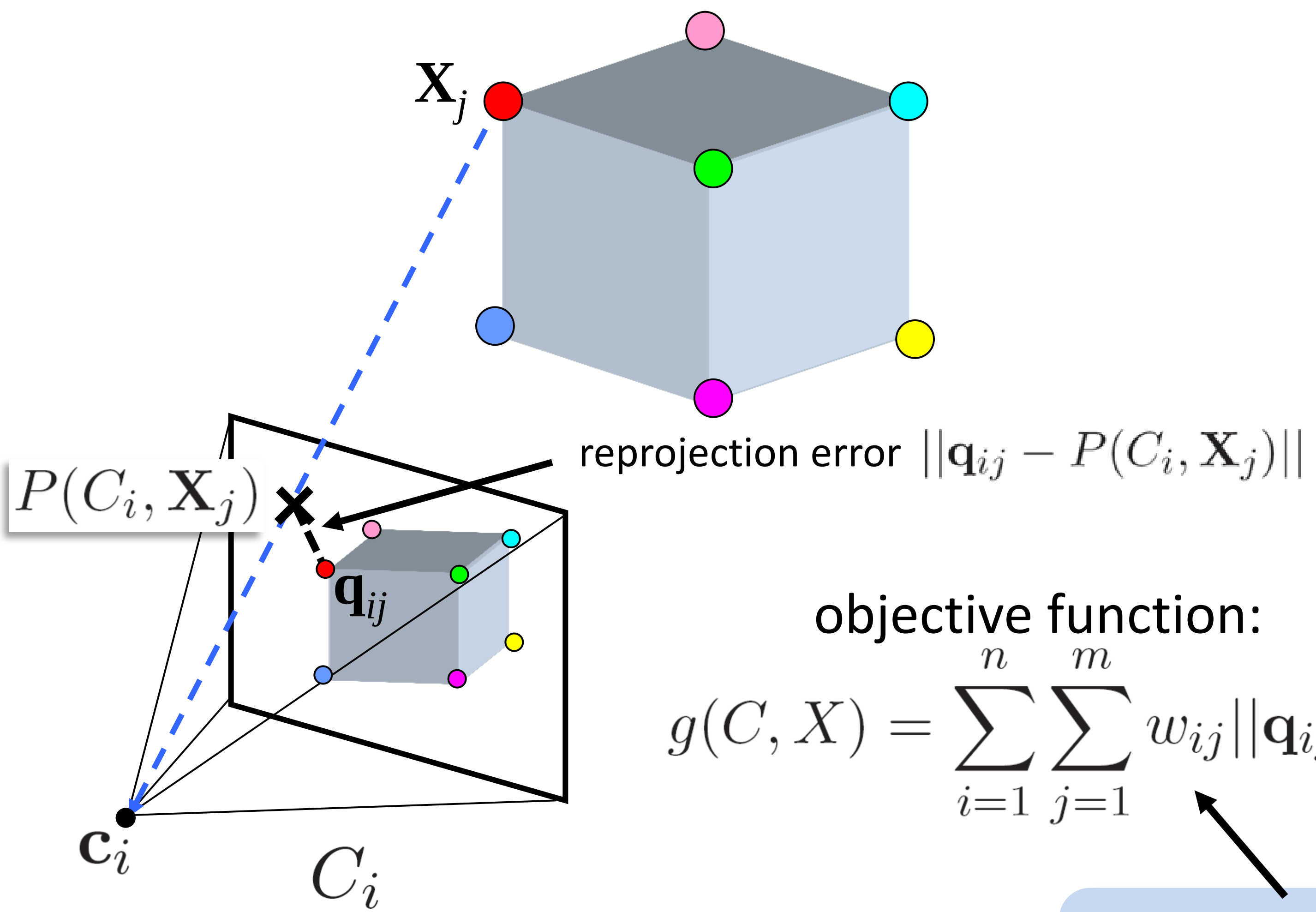
$$C = (C_1, C_2, \ldots, C_n)$$
$$X = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m)$$

  - A set of observed point projections

$\mathbf{q}_{ij}$ – the observed 2D location of point *j* in image *i*

adjust the cameras and points to minimize g, the sum of squared reprojection errors

Reprojection error



objective function:

$$g(C, X) = \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij} ||\mathbf{q}_{ij} - P(C_i, \mathbf{X}_j)||^2$$

indicator variable:
- 1 if point *j* is visible in camera *i*
- 0 otherwise

Bundle adjustment

- Minimizing *g(C,X)* is a sparse non-linear least squares problem
- This algorithm is known as *Gauss-Newton*
- In practice, a modified algorithm known as *Levenberg-Marquardt* is used

Bundle adjustment packages

- **The Ceres-Solver from Google**

  http://code.google.com/p/ceres-solver/

- **Sparse Bundle Adjustment (SBA)**

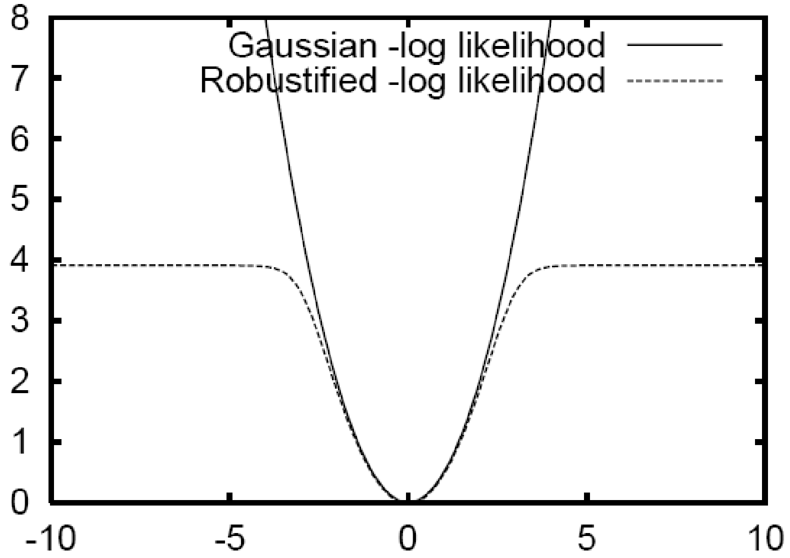
  http://www.ics.forth.gr/~lourakis/sba/

The problem of outliers

- In spite of our best efforts to get clean matches, outliers remain
- The sum-of-squared residuals objective function is statistically correct given a Gaussian noise model
  - Unfortunately, outliers break the Gaussian assumption

The problem of outliers

- Possible solutions:
  1. After each run of bundle adjustment, remove outliers and rerun
  2. Use a robust objective function



Credit: Triggs, *et al.* “Bundle adjustment – a modern synthesis”

Result: Colosseum

Result: Sanmarco

总结—本节课内容回顾

- 三角化：已知$\mathbf{x}$、$\mathbf{K}$、$\mathbf{R}$、$\mathbf{t}$，求$\mathbf{X}$
- 相机标定：已知$\mathbf{x}$、$\mathbf{X}$，求$\mathbf{K}$、$\mathbf{R}$、$\mathbf{t}$
- 稀疏重建：已知$\mathbf{x}$，求$\mathbf{K}$、$\mathbf{R}$、$\mathbf{t}$、$\mathbf{X}$
- 重投影误差最小化 $\min \sum_i \left\| \mathbf{P}\mathbf{X}_i - \mathbf{x}_i \right\|^2$
- 通过线性方法求解初始值（代数误差最小化）,通过非线性优化迭代求精（几何误差最小化）
- 相机标定与稀疏重建实践：Photo Tourism

总结—参考文献

- Multiple View Geometry in Computer Vision
  Richard Hartley and Andrew Zisserman
  Cambridge University Press, 2004.

- Bundle Adjustment: A Modern Synthesis
  B. Triggs, P. MacLauchlan, R. Hartley, A. Fitzgibbon
  ECCV 2000.

- Modeling the World from Internet Photo Collections
  N. Snavely, S. Seitz, R. Szeliski, IJCV 2008.

下节课内容——立体视觉与三维建模