

第13章 3D图像分析

中国科学技术大学
电子工程与信息科学系

主讲教师：李厚强 (lihq@ustc.edu.cn)
周文罡 (zhwg@ustc.edu.cn)
李 礼 (lill@ustc.edu.cn)
胡 洋 (eyehu@ustc.edu.cn)



3D图像分析

□ 传统的三维重建方法

- 成像变换和相机标定
- 立体视觉和对极几何
- 单应性
- 运动推断结构

□ 基于深度学习的三维重建

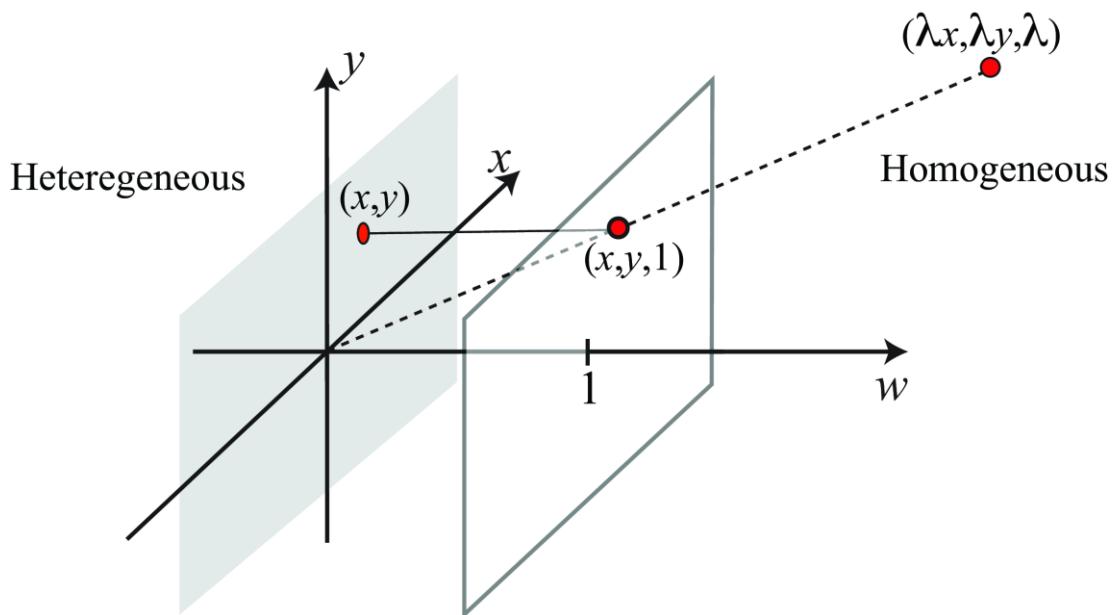
- 基于学习的单目深度估计
- 基于体素的三维表示
- 基于点云的三维表示
- 基于多边形网格的三维表示
- 基于隐函数的三维表示

成像变换

□ 齐次坐标

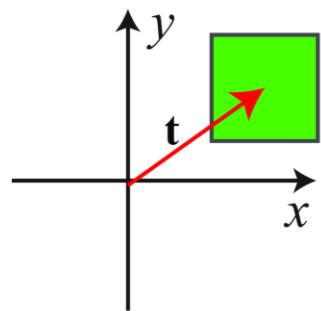
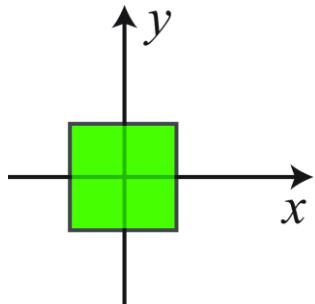
■ 笛卡尔坐标 \leftrightarrow 齐次坐标

$$\begin{pmatrix} x \\ y \end{pmatrix} \rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \equiv \begin{bmatrix} \lambda x \\ \lambda y \\ \lambda \end{bmatrix} \quad \begin{bmatrix} x \\ y \\ w \end{bmatrix} \rightarrow \begin{pmatrix} x/w \\ y/w \end{pmatrix}$$

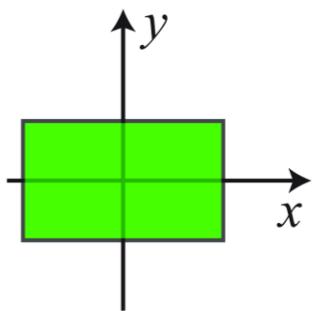


成像变换

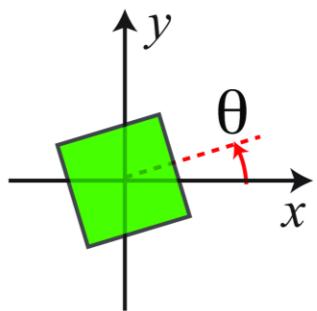
□ 基本坐标变换



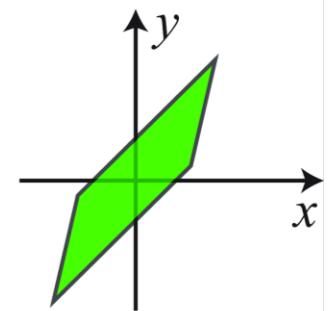
Translation



Scaling



Rotation



Shear

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

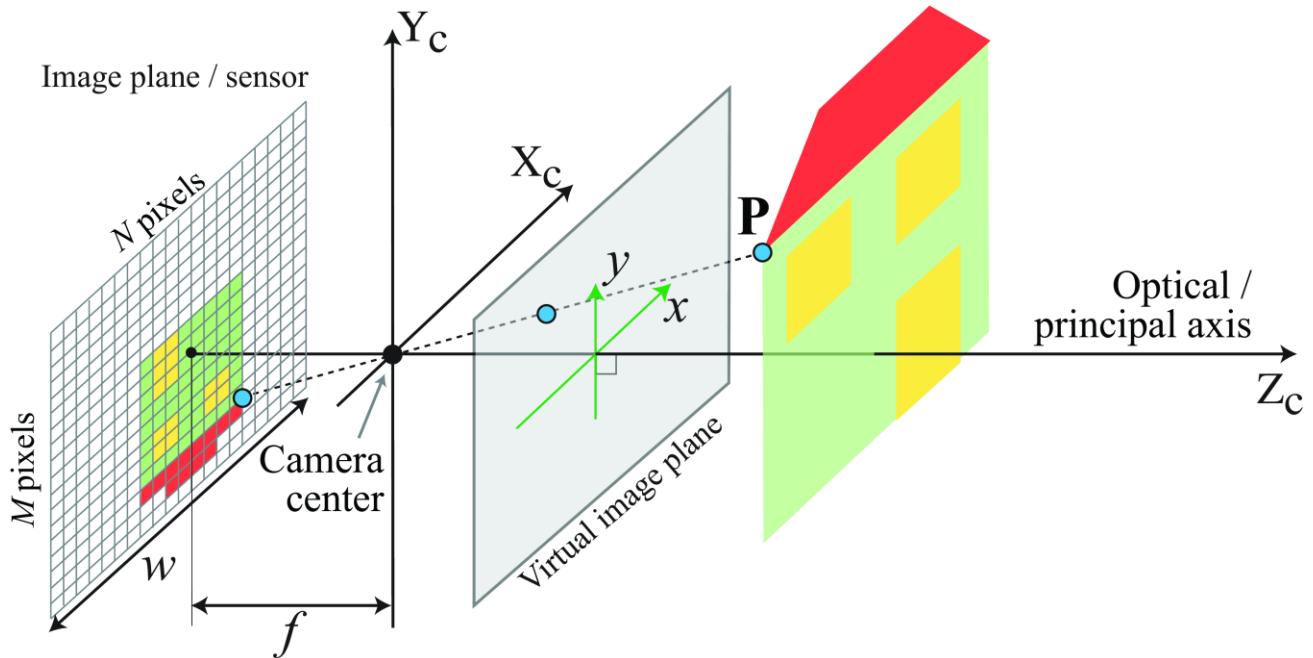
$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & q_x & 0 \\ q_y & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

成像变换



□ 相机成像模型

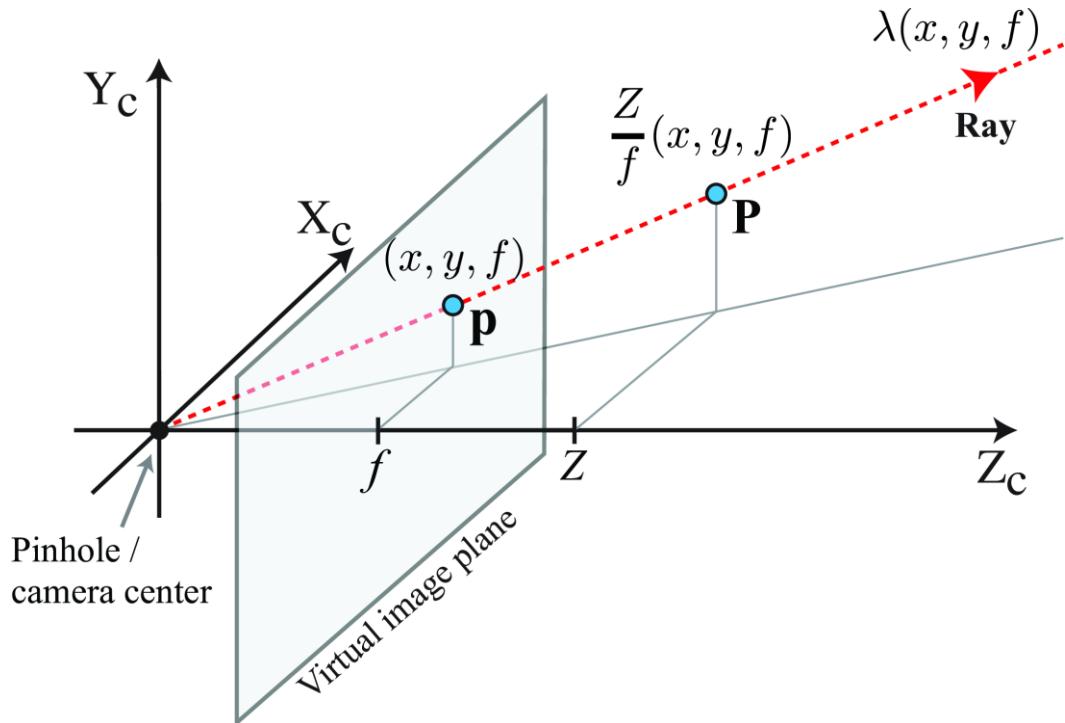


$$\begin{bmatrix} a & 0 & c_x & 0 \\ 0 & b & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \rightarrow \begin{pmatrix} aX/Z + c_x \\ bY/Z + c_y \\ 1 \end{pmatrix} = \begin{pmatrix} n \\ m \\ 1 \end{pmatrix} \quad a = fN/w$$

相机内参矩阵

成像变换

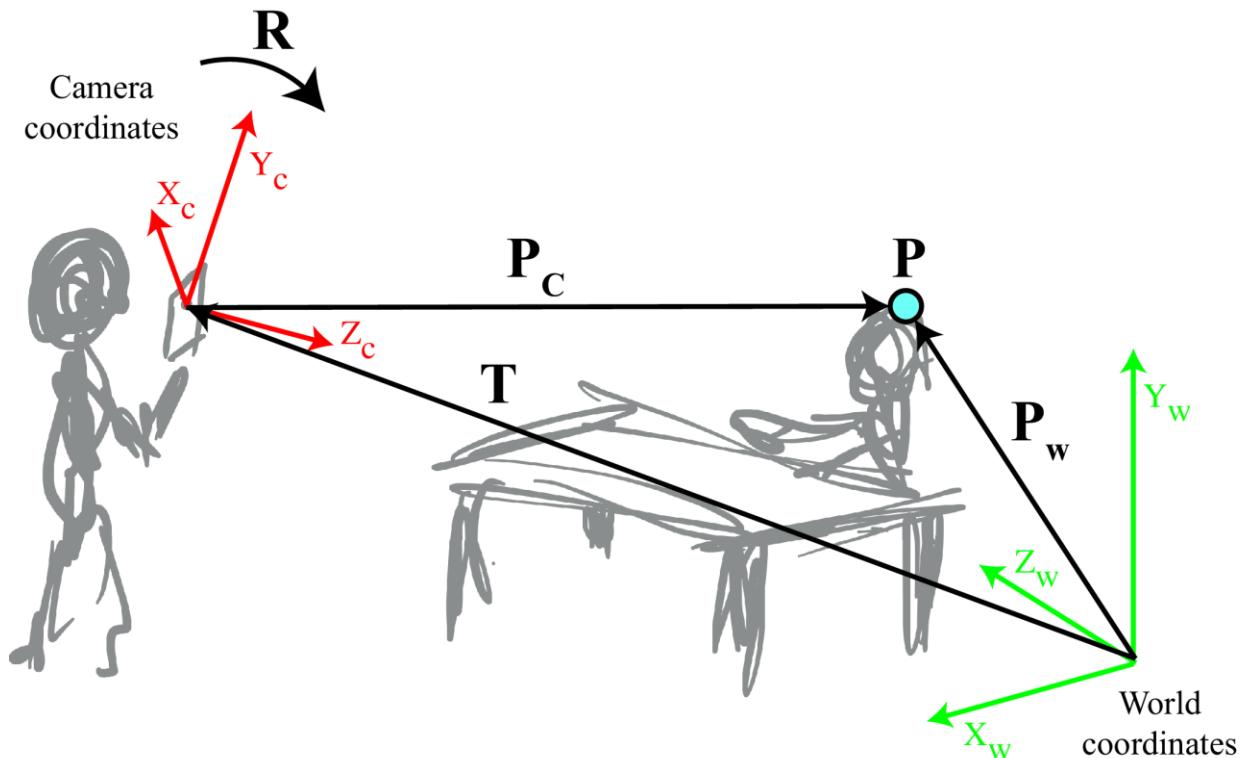
□ 逆投影（从像素到射线）



$$\frac{Z}{f} \begin{pmatrix} x \\ y \\ f \end{pmatrix} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

成像变换

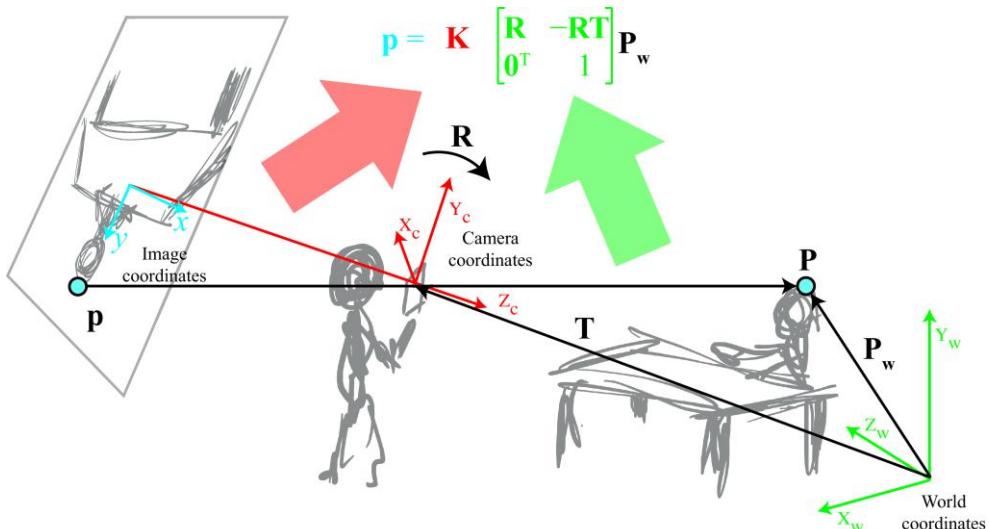
□ 相机外参矩阵：相机坐标和世界坐标不在同一位置



$$\mathbf{P}_C = \mathbf{R}(\mathbf{P}_W - \mathbf{T}) = \mathbf{R}\mathbf{P}_W - \mathbf{R}\mathbf{T} \text{ 或 } \mathbf{P}_C = \begin{bmatrix} \mathbf{R} & -\mathbf{R}\mathbf{T} \\ \mathbf{0}^T & 1 \end{bmatrix} \mathbf{P}_W$$

成像变换

□ 完整成像变换（结合外参和内参矩阵）



$$\begin{bmatrix} x' \\ y' \\ w \end{bmatrix} = \begin{bmatrix} a & 0 & c_x \\ 0 & b & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_1 & R_2 & R_3 \\ R_4 & R_5 & R_6 \\ R_7 & R_8 & R_9 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -T_X \\ 0 & 1 & 0 & -T_Y \\ 0 & 0 & 1 & -T_Z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

K是上三角矩阵；
R是正交矩阵

或 $\mathbf{p} = \mathbf{K}[\mathbf{R} \mid -\mathbf{R}\mathbf{T}]\mathbf{P}_W \quad \mathbf{M} = \mathbf{K} \begin{bmatrix} \mathbf{R} & -\mathbf{R}\mathbf{T} \\ \mathbf{0}^T & 1 \end{bmatrix}$

相机矩阵

相机标定

□ DLT(Direct Linear Transform)算法

■ 对像素 \mathbf{p}_i 和其对应的3D点 \mathbf{P}_i

$$\mathbf{p}_i = \mathbf{M}\mathbf{P}_i = \begin{bmatrix} \mathbf{m}_0^\top \\ \mathbf{m}_1^\top \\ \mathbf{m}_2^\top \end{bmatrix} \mathbf{P}_i \quad \text{即} \quad \begin{aligned} x_i &= \frac{\mathbf{m}_0^\top \mathbf{P}_i}{\mathbf{m}_2^\top \mathbf{P}_i} \\ y_i &= \frac{\mathbf{m}_1^\top \mathbf{P}_i}{\mathbf{m}_2^\top \mathbf{P}_i} \end{aligned}$$

整理成矩阵形式

$$\begin{bmatrix} -\mathbf{P}_i^\top & \mathbf{0} & x_i \mathbf{P}_i^\top \\ \mathbf{0} & -\mathbf{P}_i^\top & y_i \mathbf{P}_i^\top \end{bmatrix} \begin{bmatrix} \mathbf{m}_0 \\ \mathbf{m}_1 \\ \mathbf{m}_2 \end{bmatrix} = \mathbf{0}$$

■ 对N对点

$$\begin{bmatrix} -\mathbf{P}_1^\top & \mathbf{0} & x_1 \mathbf{P}_1^\top \\ \mathbf{0} & -\mathbf{P}_1^\top & y_1 \mathbf{P}_1^\top \\ \vdots & \vdots & \vdots \\ -\mathbf{P}_N^\top & \mathbf{0} & x_N \mathbf{P}_N^\top \\ \mathbf{0} & -\mathbf{P}_N^\top & y_N \mathbf{P}_N^\top \end{bmatrix} \begin{bmatrix} \mathbf{m}_0 \\ \mathbf{m}_1 \\ \mathbf{m}_2 \end{bmatrix} = \mathbf{0} \quad \text{可解得M}$$



相机标定

□ 由M恢复相机内参和外参

$$\mathbf{M} = [\mathbf{KR} \mid -\mathbf{KRT}]$$

■ 计算T

设 $\mathbf{M} = [\mathbf{B} \ \mathbf{b}]$, $\mathbf{B} = \mathbf{KR}$, $\mathbf{b} = -\mathbf{KRT}$, 则

$$\mathbf{T} = -\mathbf{B}^{-1}\mathbf{b}$$

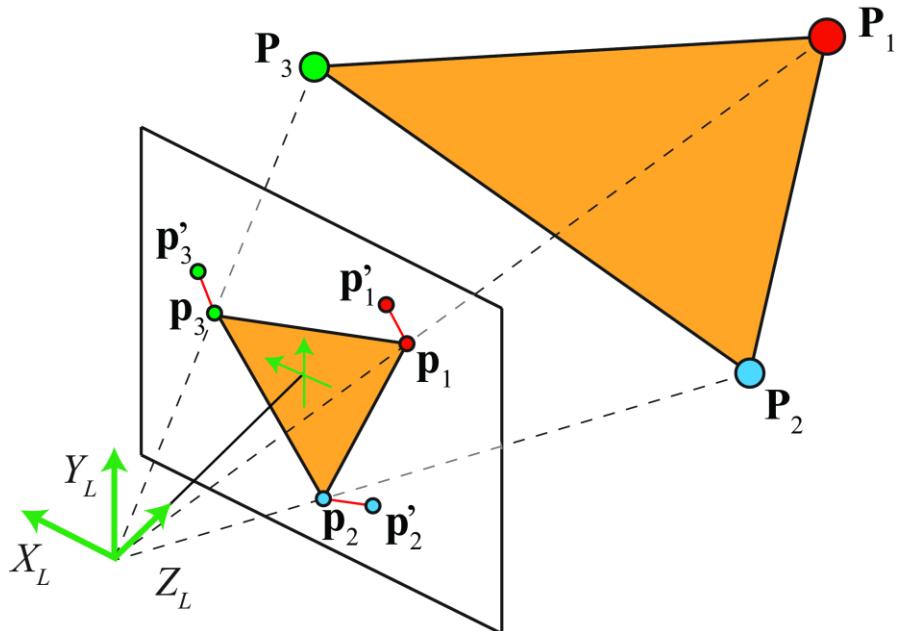
■ 计算K, R

由于K为上三角矩阵, R为正交矩阵, 通过对B⁻¹进行QR分解可得到对K, R的估计

■ 此估计算法对噪声较敏感, 但得到的相机参数可作为其他标定算法的初始值

相机标定

□ 最小重投影误差法



$$\sum_{i=1}^N \|\mathbf{p}_i - \mathbf{p}'_i\|^2 = \sum_{i=1}^N \|\mathbf{p}_i - \pi(\mathbf{K}[\mathbf{R} \mid - \mathbf{R}\mathbf{T}]\mathbf{P}_i)\|^2$$

Π函数将齐次坐标变为笛卡尔坐标



3D图像分析

□ 传统的三维重建方法

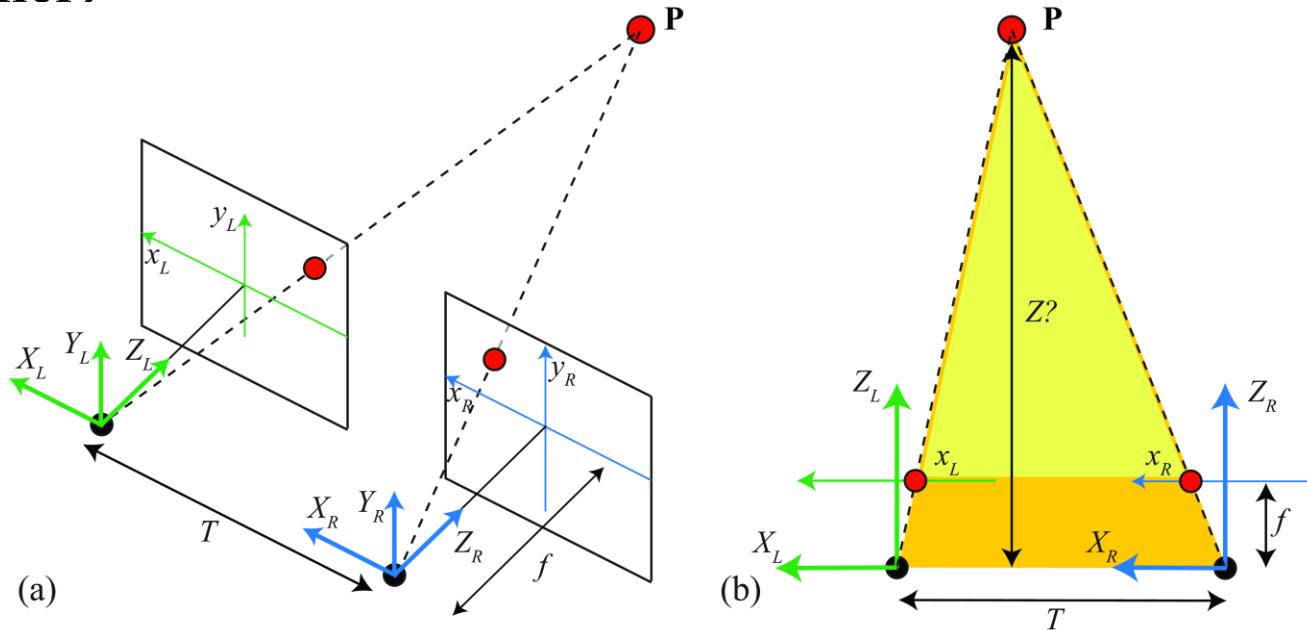
- 成像变换和相机标定
- 立体视觉和对极几何
- 单应性
- 运动推断结构

□ 基于深度学习的三维重建

- 基于学习的单目深度估计
- 基于体素的三维表示
- 基于点云的三维表示
- 基于多边形网格的三维表示
- 基于隐函数的三维表示

立体视觉

- 立体视觉原理 (both cameras are identical (same intrinsic parameters), and one is translated horizontally with respect to the other)



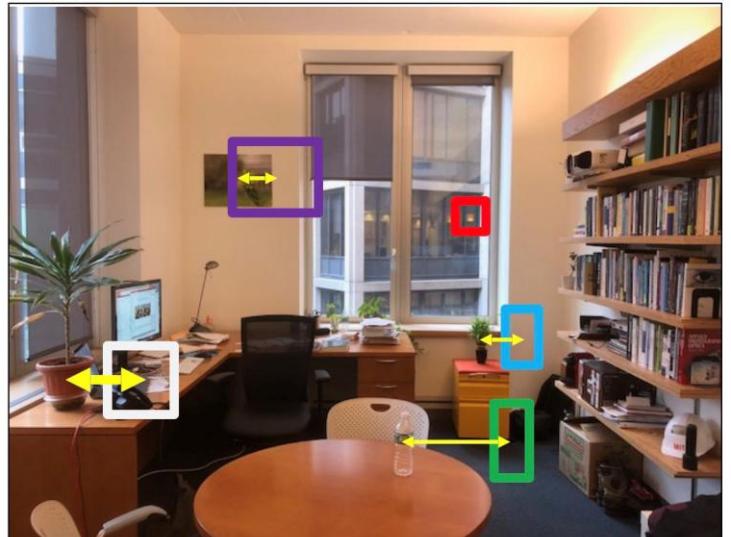
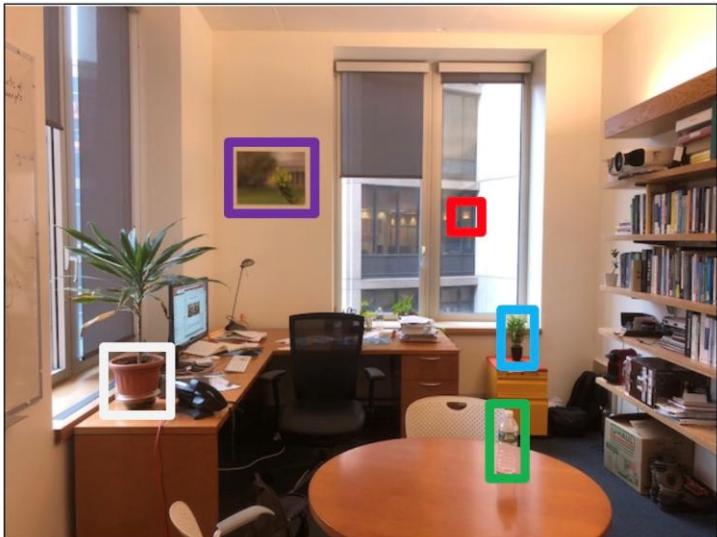
$$\frac{T + x_L - x_R}{Z - f} = \frac{T}{Z} \quad \text{即} \quad Z = \frac{fT}{x_R - x_L}$$

视差 (Disparity)

立体视觉

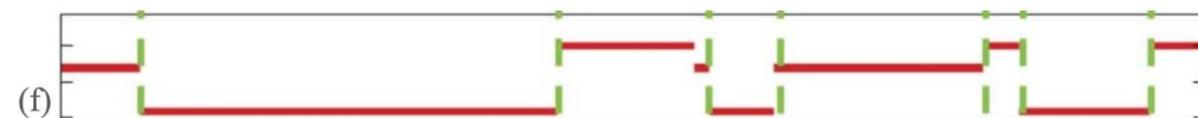
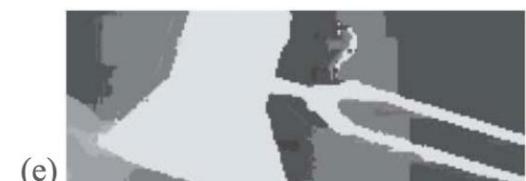
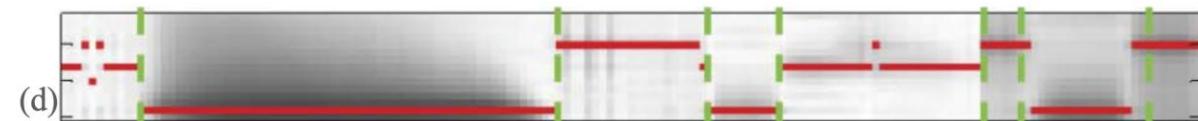
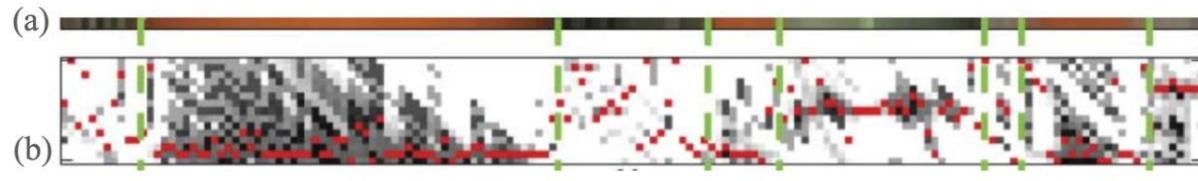
□ 立体匹配

- 不清楚两个视频的对应点



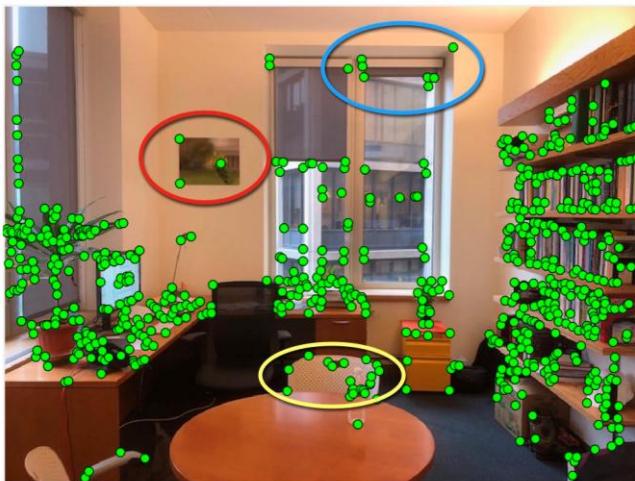
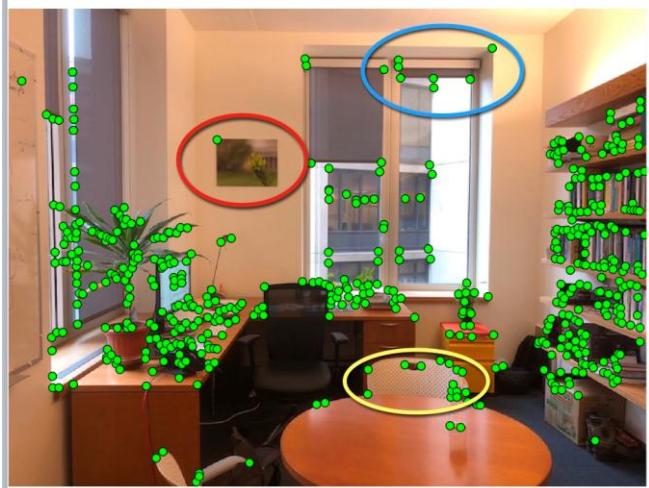
立体视觉

□ 基于亮度的匹配



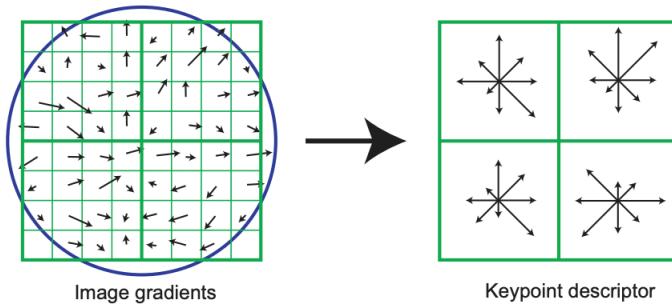
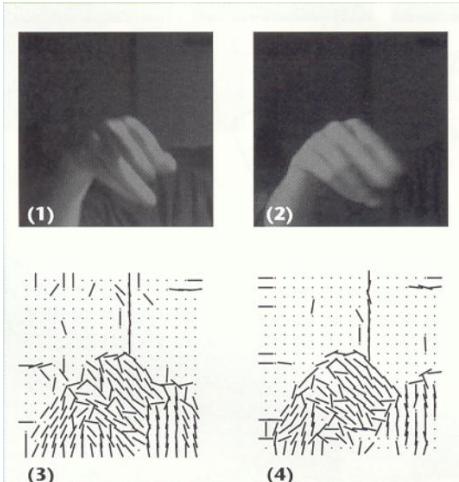
立体视觉

- 基于局部特征的匹配
 - 关键点检测（比如Harris角点）

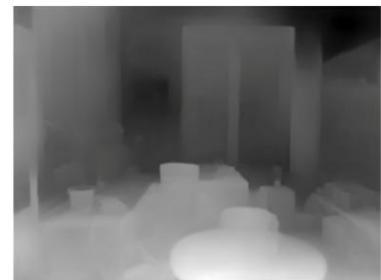
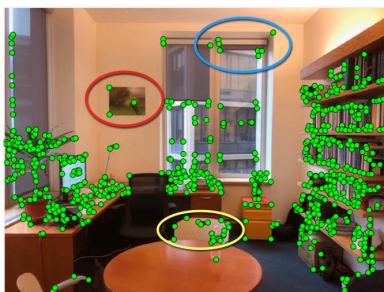


立体视觉

- 基于局部特征的匹配
- 局部图像描述 (SIFT)

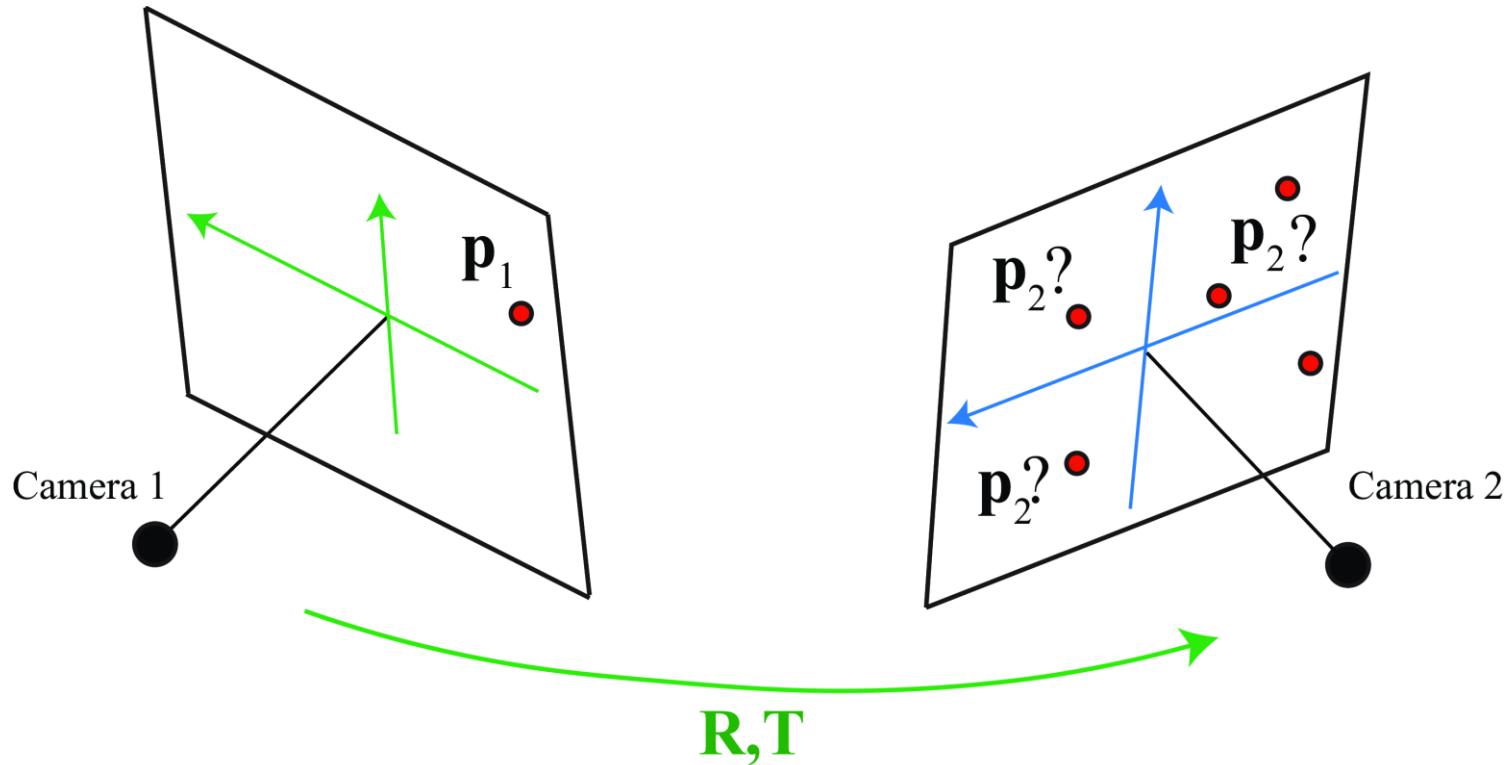


- 深度插值



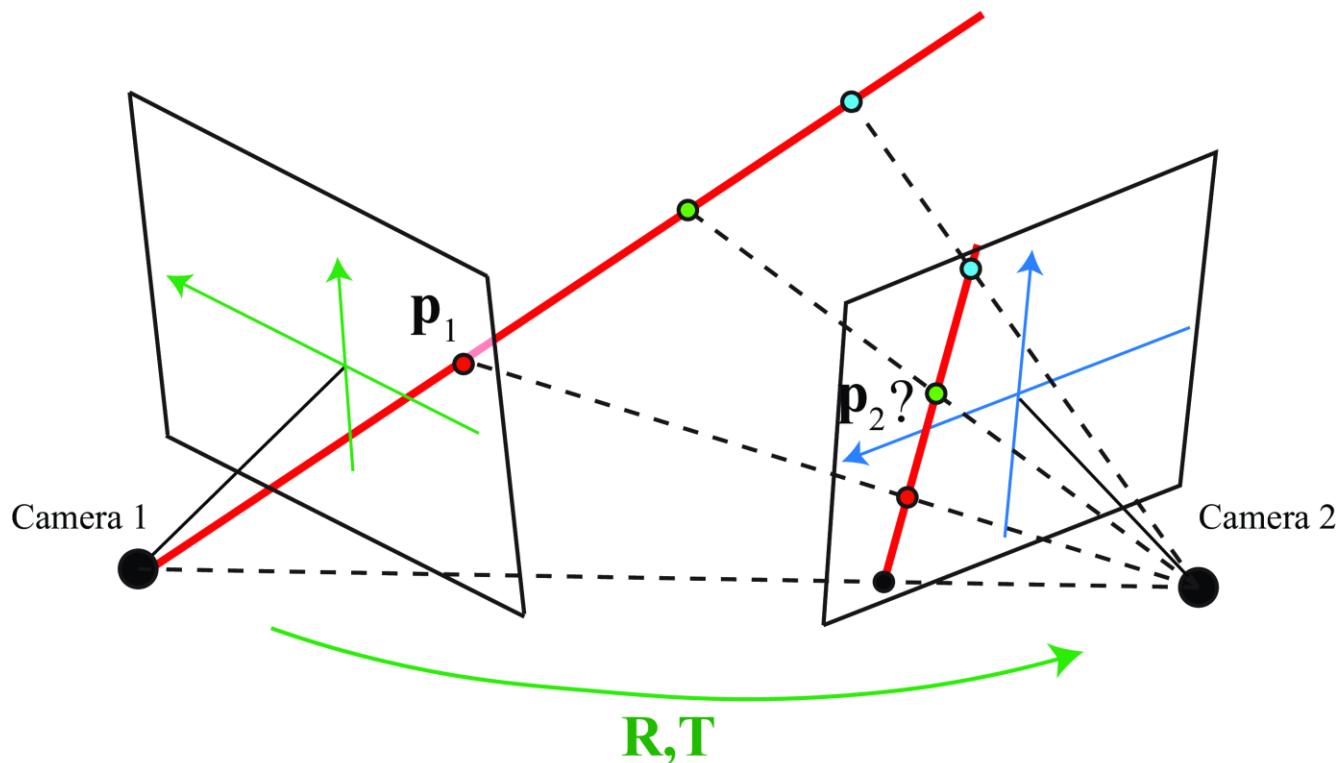
对极几何

- 描述两个不同视图下，一对匹配点之间的几何约束关系
(两个视角不在一个水平线的情况)



对极几何

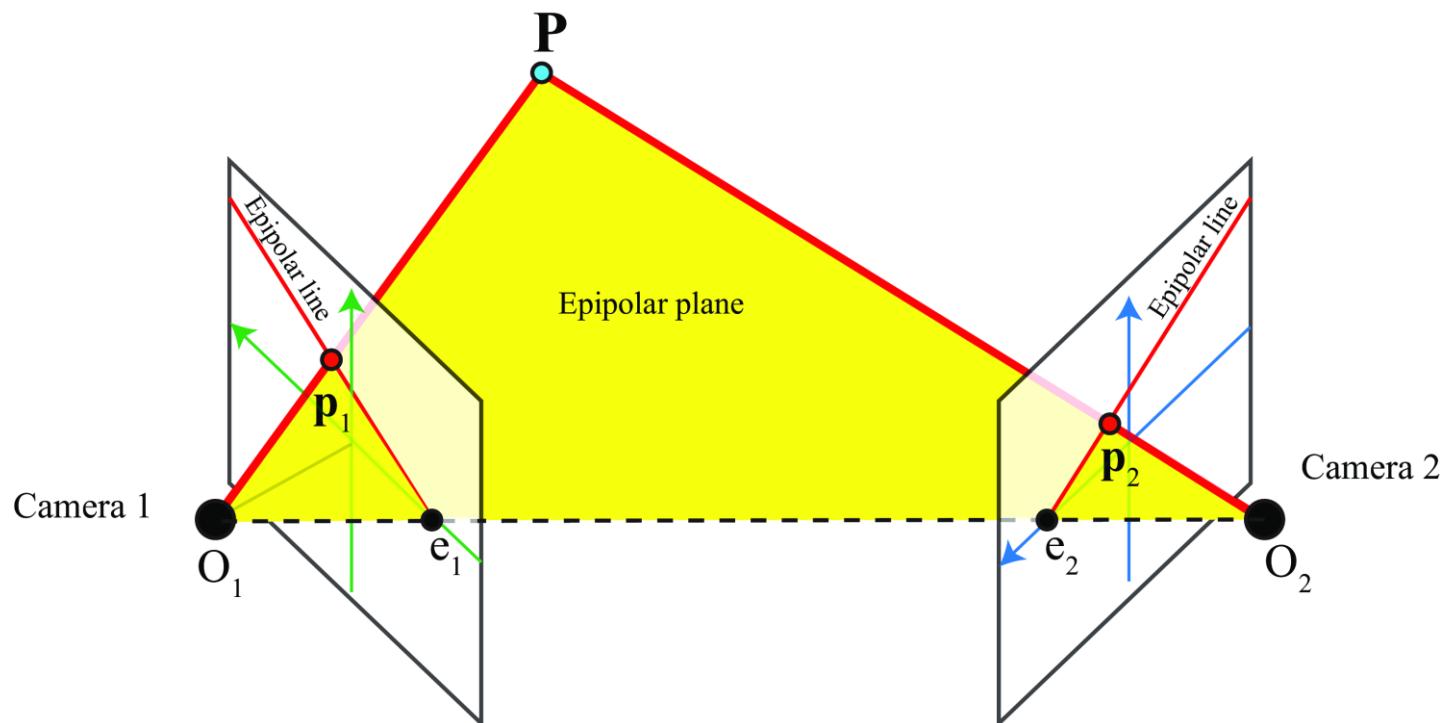
- 描述两个不同视图下，一对匹配点之间的几何约束关系
(两个视角不在一个水平线的情况)



对极几何

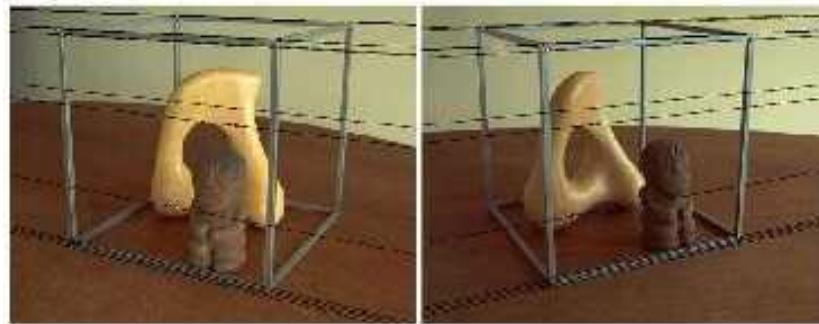
□ 描述两个不同视图下，一对匹配点之间的几何约束关系

- Epipolar plane
- Epipolar line
- Epipoles



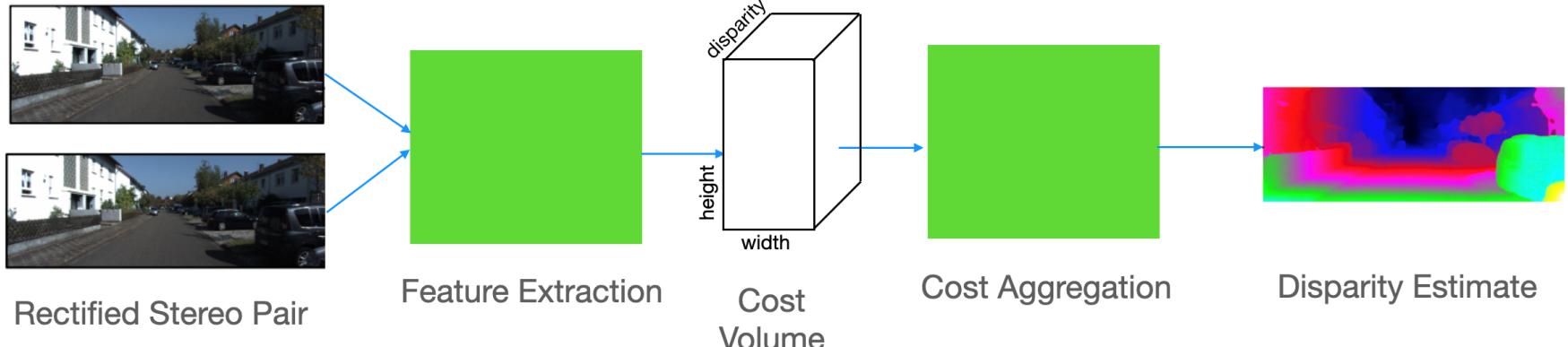
图像矫正(Image Rectification)

- 对任意视角获取的两张图，通过对相机进行旋转、并对图像进行缩放等变换，使得两张图像位于同一水平高度且相互平行
- 矫正后，两张图像上的对应点位于具有相同高度点水平扫描线上，从而可简化特征匹配



基于学习的方法

□ 基于学习的深度估计



$$J(\theta) = \sum_t |\mathbf{d} - f_{\theta}(\mathbf{v}_1, \mathbf{v}_2)|^2$$

$$f(\mathbf{v}_1, \mathbf{v}_2) = c(g(\mathbf{v}_1), g(\mathbf{v}_2))$$

J.-R. Chang, Y.-S. Chen, Pyramid stereo matching network., in: Proceedings of the IEEE/CVF Computer Vision and Pattern Recognition, 2018.



3D图像分析

□ 传统的三维重建方法

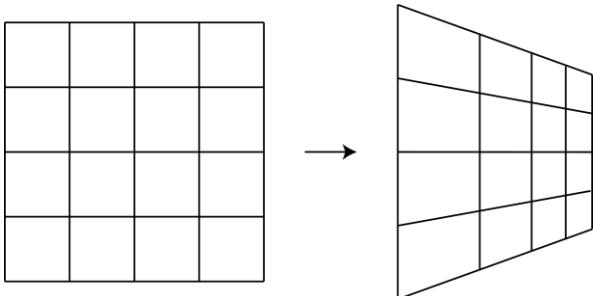
- 成像变换和相机标定
- 立体视觉和对极几何
- **单应性**
- 运动推断结构

□ 基于深度学习的三维重建

- 基于学习的单目深度估计
- 基于体素的三维表示
- 基于点云的三维表示
- 基于多边形网格的三维表示
- 基于隐函数的三维表示

单应性(Homography)

- 单应性变换：点变换函数 $p' = h(p)$ ，若点 p_1, p_2, p_3 在一条直线上，则变换后的点 $h(p_1), h(p_2), h(p_3)$ 也共线



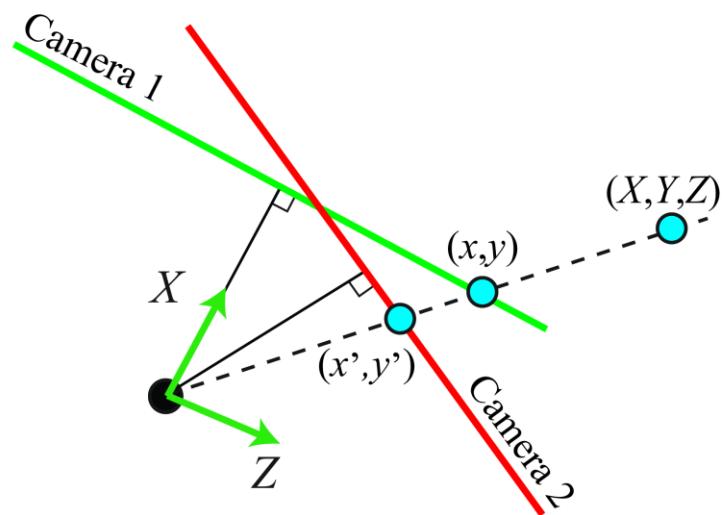
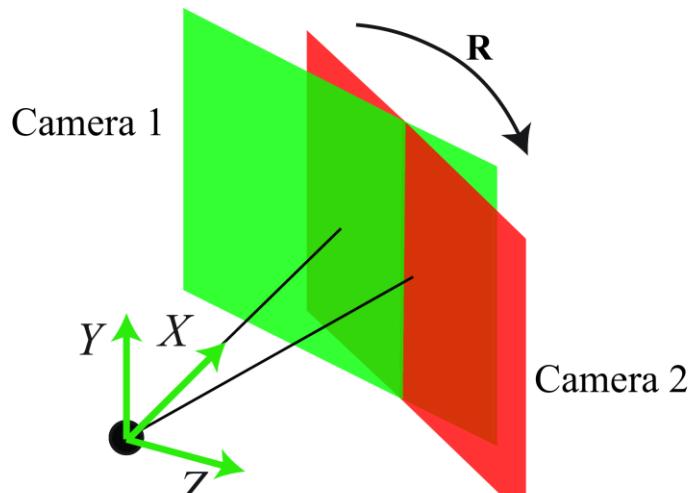
- 用齐次坐标可表示为

$$\mathbf{p}' = \mathbf{H}\mathbf{p}$$

- 3×3 矩阵 \mathbf{H} 为单应矩阵，8个自由度
- 共线证明：对直线 $\mathbf{l}^\top \mathbf{p} = 0$ ，又 $\mathbf{p} = \mathbf{H}^{-1}\mathbf{p}'$ ，则有 $(\mathbf{H}^{-1}\mathbf{l})^\top \mathbf{p}' = 0$ ，即变换后仍然为直线

单应性(Homography)

□ 相机仅旋转，图像存在单应性



$$\lambda_1 \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{K}[\mathbf{I} \quad \mathbf{0}] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \mathbf{K} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$\lambda_2 \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \mathbf{K}[\mathbf{R} \quad \mathbf{0}] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \mathbf{K}\mathbf{R} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

单应性(Homography)

□ 相机仅旋转，图像存在单应性

$$\lambda_1 \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{K} [\mathbf{I} \quad \mathbf{0}] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \mathbf{K} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$\lambda_2 \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \mathbf{K} [\mathbf{R} \quad \mathbf{0}] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \mathbf{K} \mathbf{R} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

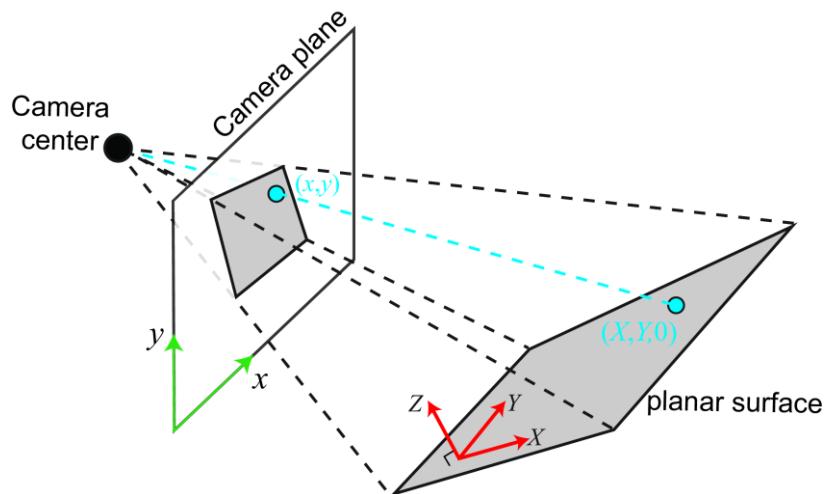
$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \lambda_1 \mathbf{K}^{-1} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \lambda_2 \mathbf{R}^{-1} \mathbf{K}^{-1} \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix}$$

$$\lambda_2 / \lambda_1 \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \mathbf{K} \mathbf{R} \mathbf{K}^{-1} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

即 $\lambda \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \mathbf{H} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$

单应性(Homography)

□ 平面物体图像存在单应性



$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{K}[\mathbf{R} \quad \mathbf{t}] \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = \mathbf{K}[\mathbf{c}_1 \quad \mathbf{c}_2 \quad \mathbf{t}] \begin{bmatrix} X_1 \\ Y_1 \\ 1 \end{bmatrix}$$

即 $\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{H} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$

单应性(Homography)

□ 应用举例：全景图像拼接

- 仅有摄像头转动得到的不同视角的图像完成拼接



单应性(Homography)

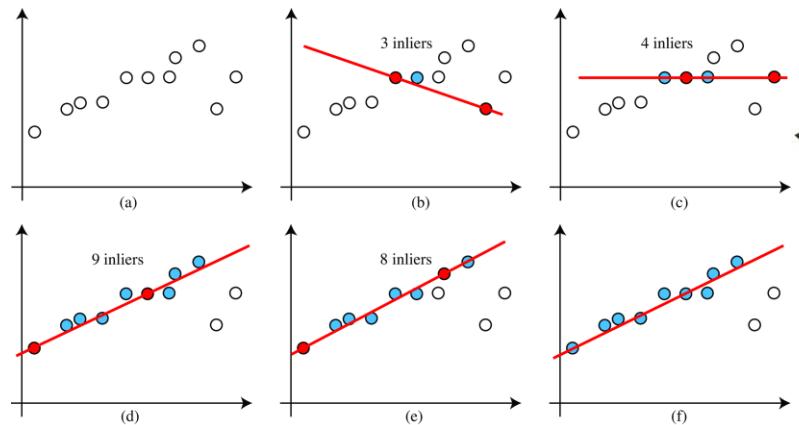
□ 应用举例：全景图像拼接

■ DLT算法

$$A\mathbf{h} = \mathbf{0}$$

- ✓ A 为 $2N \times 9$ 矩阵, \mathbf{h} 为 9×1 向量
- ✓ 最小化 $\|A\mathbf{h}\|^2$, 最优解为 A 的最小奇异值对应的奇异向量, 或 $A^T A$ 的最小特征值对应的特征向量

■ RANSAC





3D图像分析

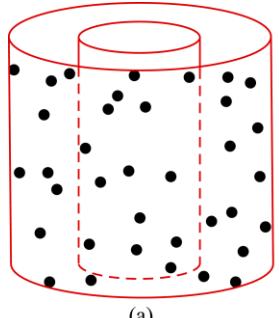
□ 传统的三维重建方法

- 成像变换和相机标定
- 立体视觉和对极几何
- 单应性
- **运动推断结构**

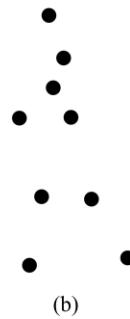
□ 基于深度学习的三维重建

- 基于学习的单目深度估计
- 基于体素的三维表示
- 基于点云的三维表示
- 基于多边形网格的三维表示
- 基于隐函数的三维表示

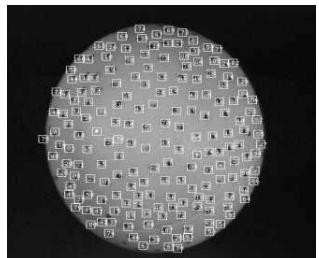
运动推断结构(Structure from Motion)



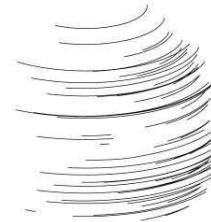
(a)



(b)



(a)



(b)



(c)



(a)

Photo Tourism
Exploring photo collections in 3D



(b)

Microsoft



(c)



运动推断结构(Structure from Motion)

□ 问题定义

- 假设场景中有 N 个不同的3D点 \mathbf{P}_i , 有 M 个不同的视角
- 对每个视角, 相机参数为 $\mathbf{K}^{(j)}, \mathbf{R}^{(j)}, \mathbf{T}^{(j)}$
- 对每个视角, 点 \mathbf{P}_i 的投影为 $\mathbf{p}_i^{(j)}$
- SFM任务定义为给定一系列有对应关系的2D点 $\mathbf{p}_i^{(j)}$, 恢复其对应的3D点 \mathbf{P}_i , 及 M 组相机参数 $\mathbf{K}^{(j)}, \mathbf{R}^{(j)}, \mathbf{T}^{(j)}$

□ 分析

- 设相机内参 $\mathbf{K}^{(j)}$ 已知, 以第一个相机确定世界坐标
- 为实现重建, 观察点数需满足

$$2NM \geq 3N + 6(M - 1) - 1$$

- 实际中有某些点在某些视角不可见的情况, 则需要更多观察点



运动推断结构(Structure from Motion)

□ 光束平差法(Bundle Adjustment)

■ 最小化投影误差

$$\sum_{j=1}^M \sum_{i=1}^N \left\| \mathbf{p}_i^{(j)} - \hat{\mathbf{p}}_i^{(j)} \right\|^2$$

$$\sum_{j=1}^M \sum_{i=1}^N \left\| \mathbf{p}_i^{(j)} - \pi \left(\mathbf{K}^{(j)} \left[\mathbf{R}^{(j)} \mid - \mathbf{R}^{(j)} \mathbf{T}^{(j)} \right] \mathbf{P}_i \right) \right\|^2$$

考虑可见性

$$\sum_{j=1}^M \sum_{i=1}^N v_i^{(j)} \left\| \mathbf{p}_i^{(j)} - \pi \left(\mathbf{K}^{(j)} \left[\mathbf{R}^{(j)} \mid - \mathbf{R}^{(j)} \mathbf{T}^{(j)} \right] \mathbf{P}_i \right) \right\|^2$$



运动推断结构(Structure from Motion)

□ 光束平差法(Bundle Adjustment)

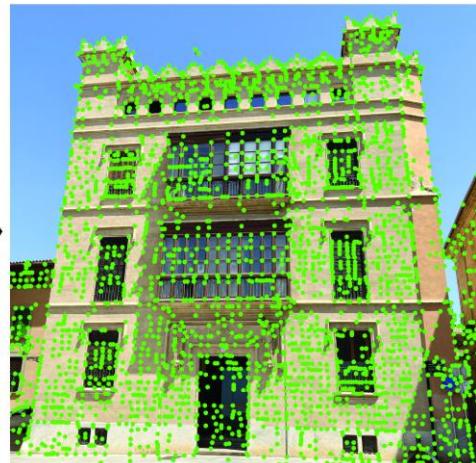
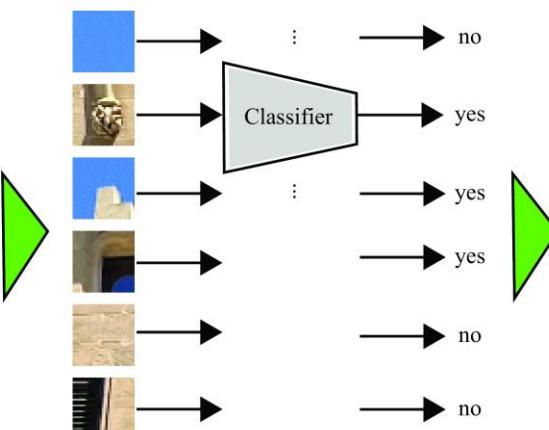
$$\sum_{j=1}^M \sum_{i=1}^N v_i^{(j)} \left\| \mathbf{p}_i^{(j)} - \pi \left(\mathbf{K}^{(j)} \left[\mathbf{R}^{(j)} \mid - \mathbf{R}^{(j)} \mathbf{T}^{(j)} \right] \mathbf{P}_i \right) \right\|^2$$

- 非线性优化问题, 局部最优
- 批量优化 (batch optimization)
- 渐进优化 (Progressive optimization)
 - ✓ 先以含匹配点最多的两张图进行重建
 - ✓ 逐渐添加新的相机

运动推断结构(Structure from Motion)

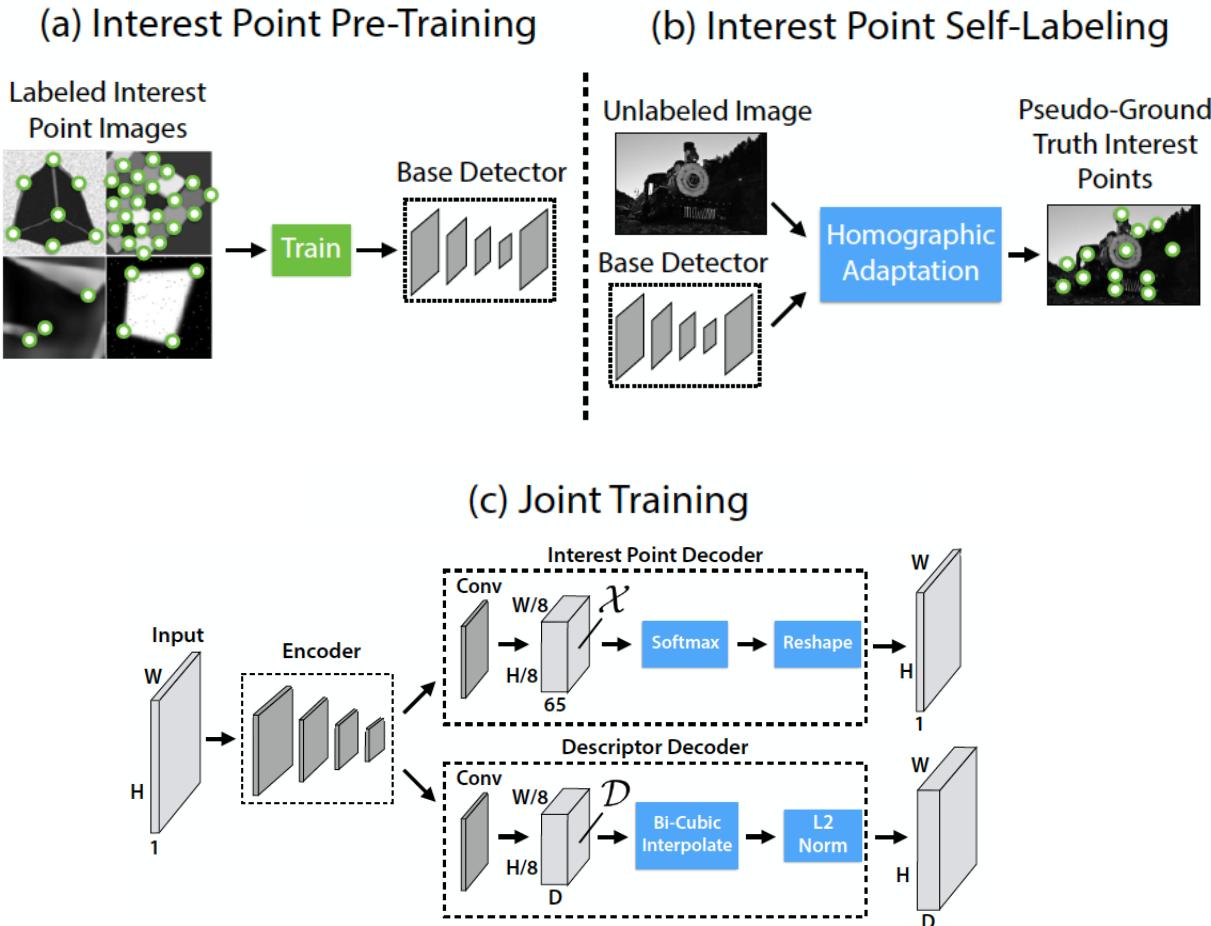
□ 基于学习的关键点检测和局部特征表示

■ SuperPoint



运动推断结构(Structure from Motion)

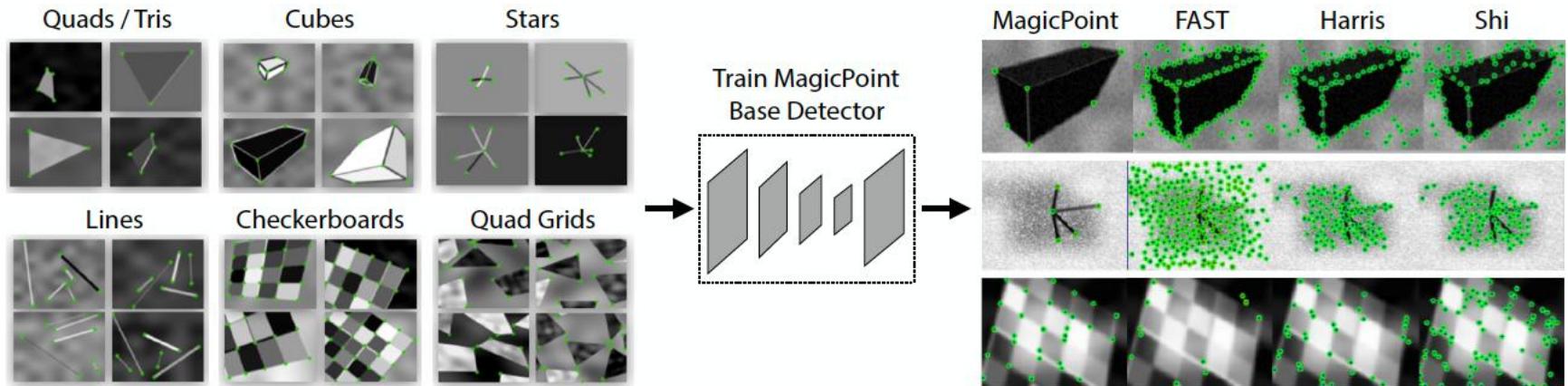
□ 基于学习的关键点检测和局部特征表示



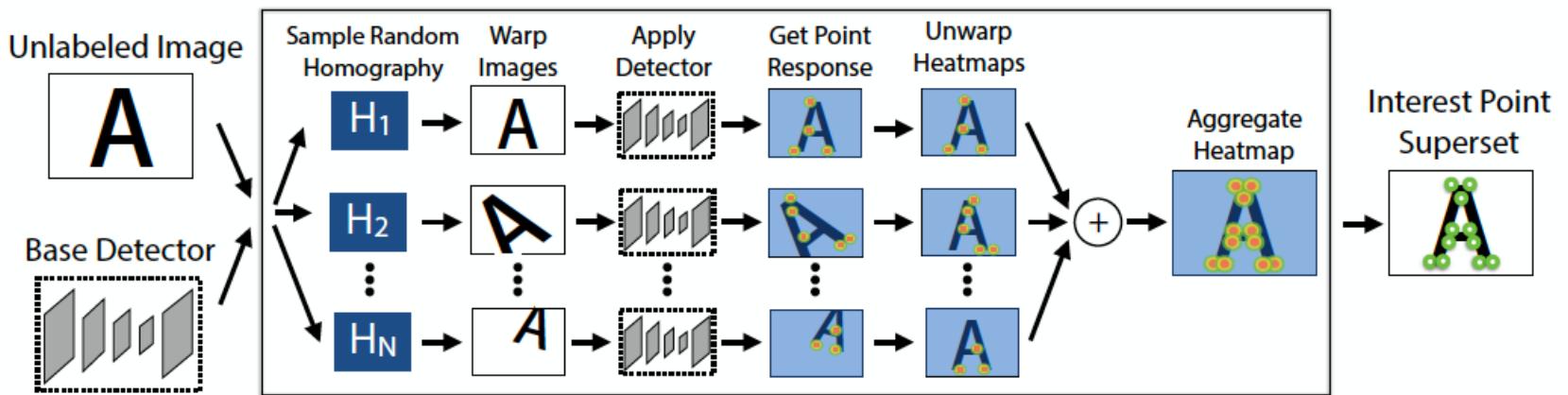
- D. DeTone, T. Malisiewicz, A. Rabinovich, SuperPoint: Self-supervised interest point detection and description. In CVPRW, 2018.

运动推断结构(Structure from Motion)

□ 基于学习的关键点检测和局部特征表示



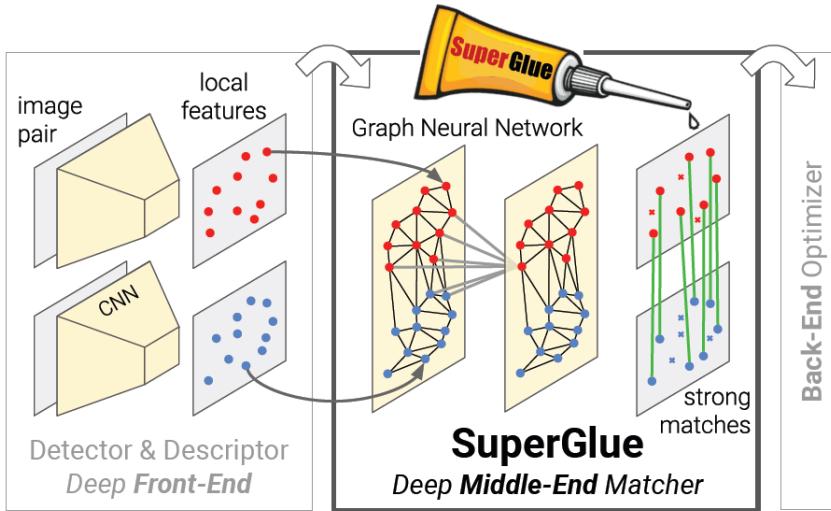
Homographic Adaptation



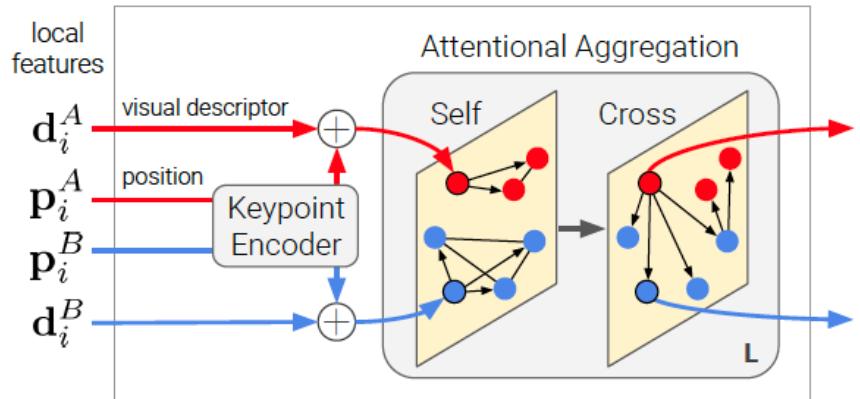
- D. DeTone, T. Malisiewicz, A. Rabinovich, SuperPoint: Self-supervised interest point detection and description. In CVPRW, 2018.

运动推断结构(Structure from Motion)

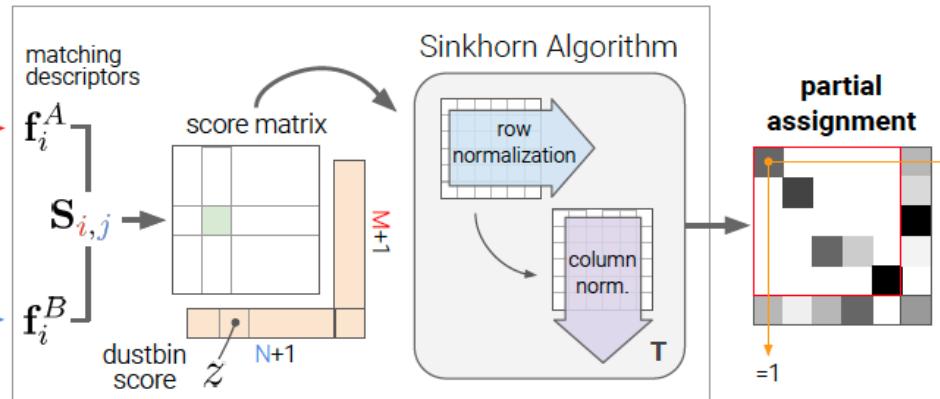
□ 基于学习的关键点匹配



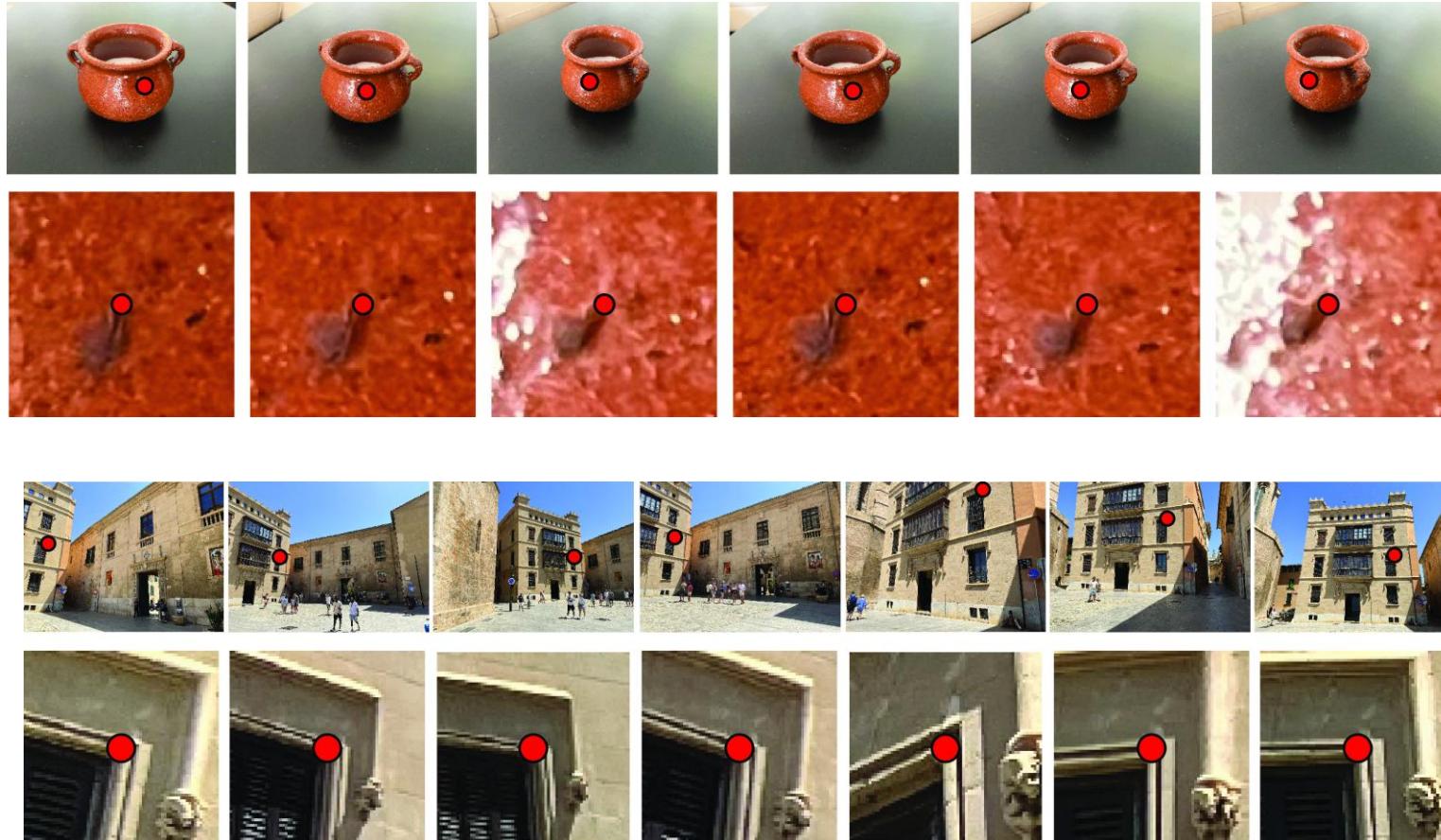
Attentional Graph Neural Network



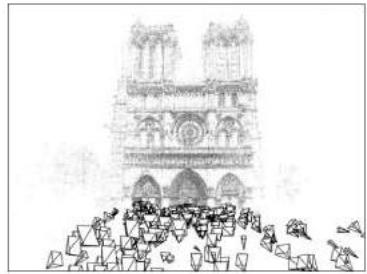
Optimal Matching Layer



运动推断结构(Structure from Motion)



运动推断结构(Structure from Motion)



- N. Snavely, S. M. Seitz, R. SzeliskiPhoto, Tourism: Exploring Photo Collections in 3D. In SIGGRAPH, 2006.



3D图像分析

□ 传统的三维重建方法

- 成像变换和相机标定
- 立体视觉和对极几何
- 单应性
- 运动推断结构

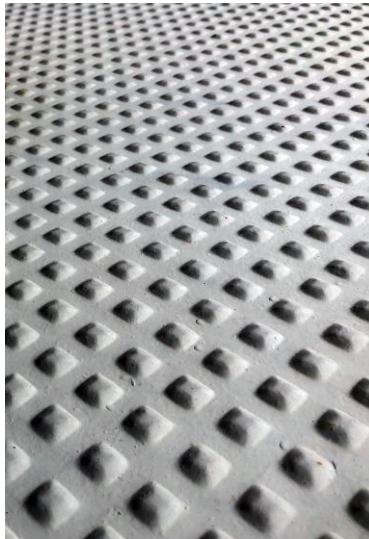
□ 基于深度学习的三维重建

- 基于学习的单目深度估计
- 基于体素的三维表示
- 基于点云的三维表示
- 基于多边形网格的三维表示
- 基于隐函数的三维表示

单目深度估计

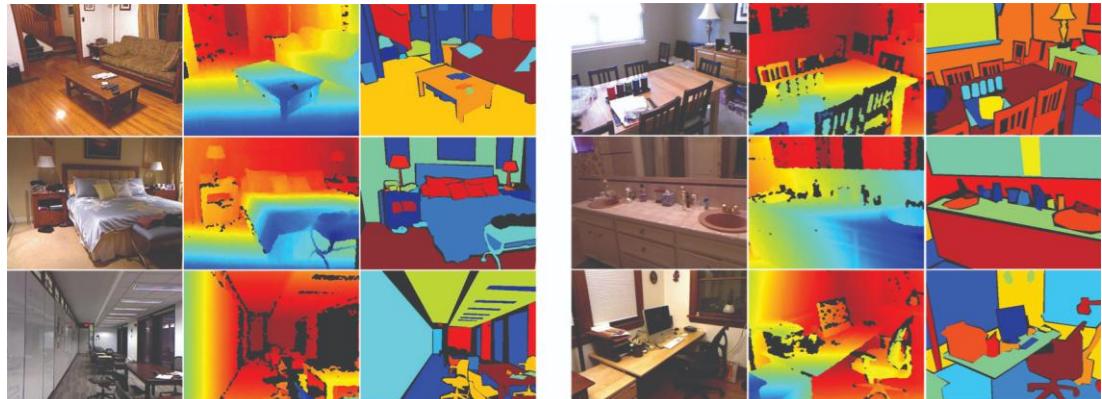
□ 图像中的深度线索 (So-called prior)

- 影调(Shading): 表面朝向不同引起的亮度的空间变化
- 纹理元尺寸、形状、空间关系等变化
- 阴影和相互反射
- 大气透视(Atmospheric perspective): 大气中光线的散射和吸收, 使图像中的物体随距离增加而颜色变淡、对比减小、模糊
- 相似物体

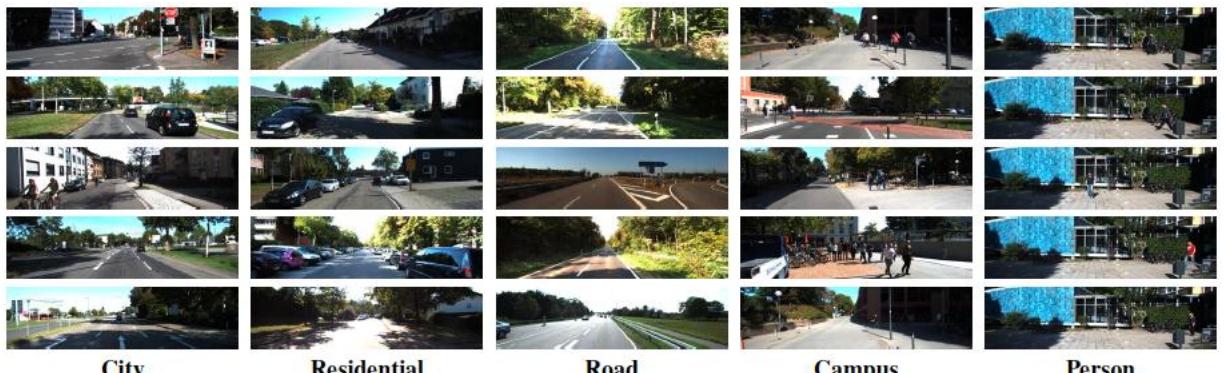
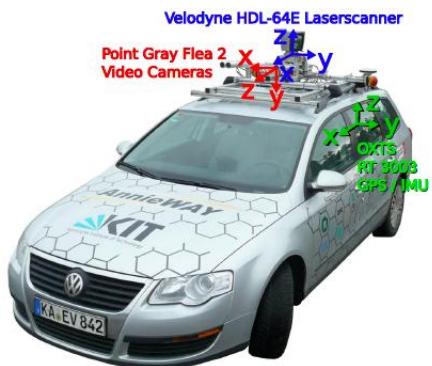


基于学习的单目深度估计

□ 数据采集



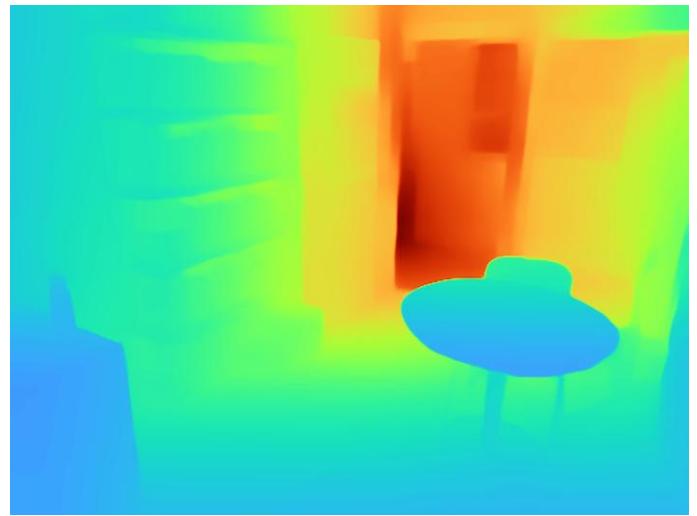
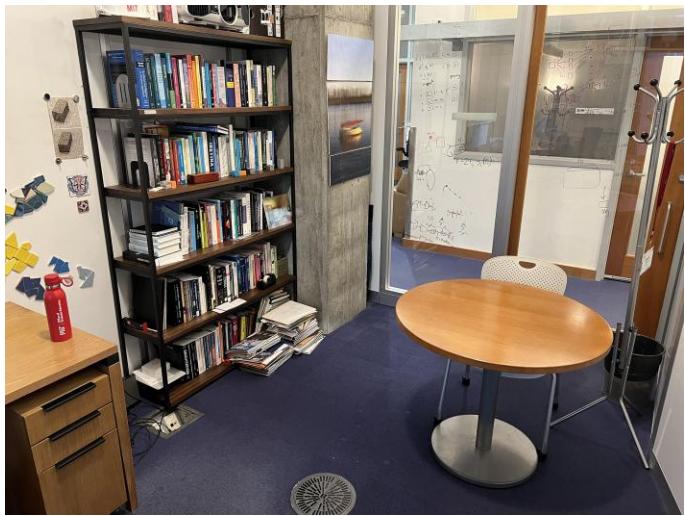
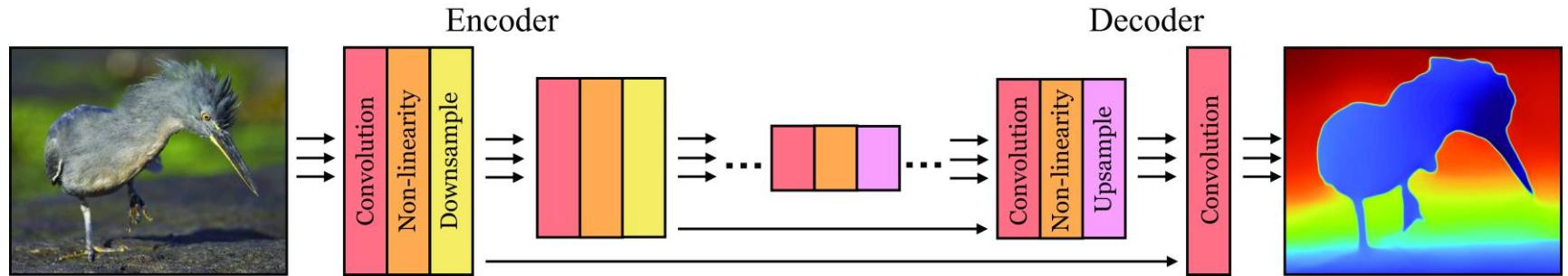
NYUv2 dataset



KITTI dataset

基于学习的单目深度估计

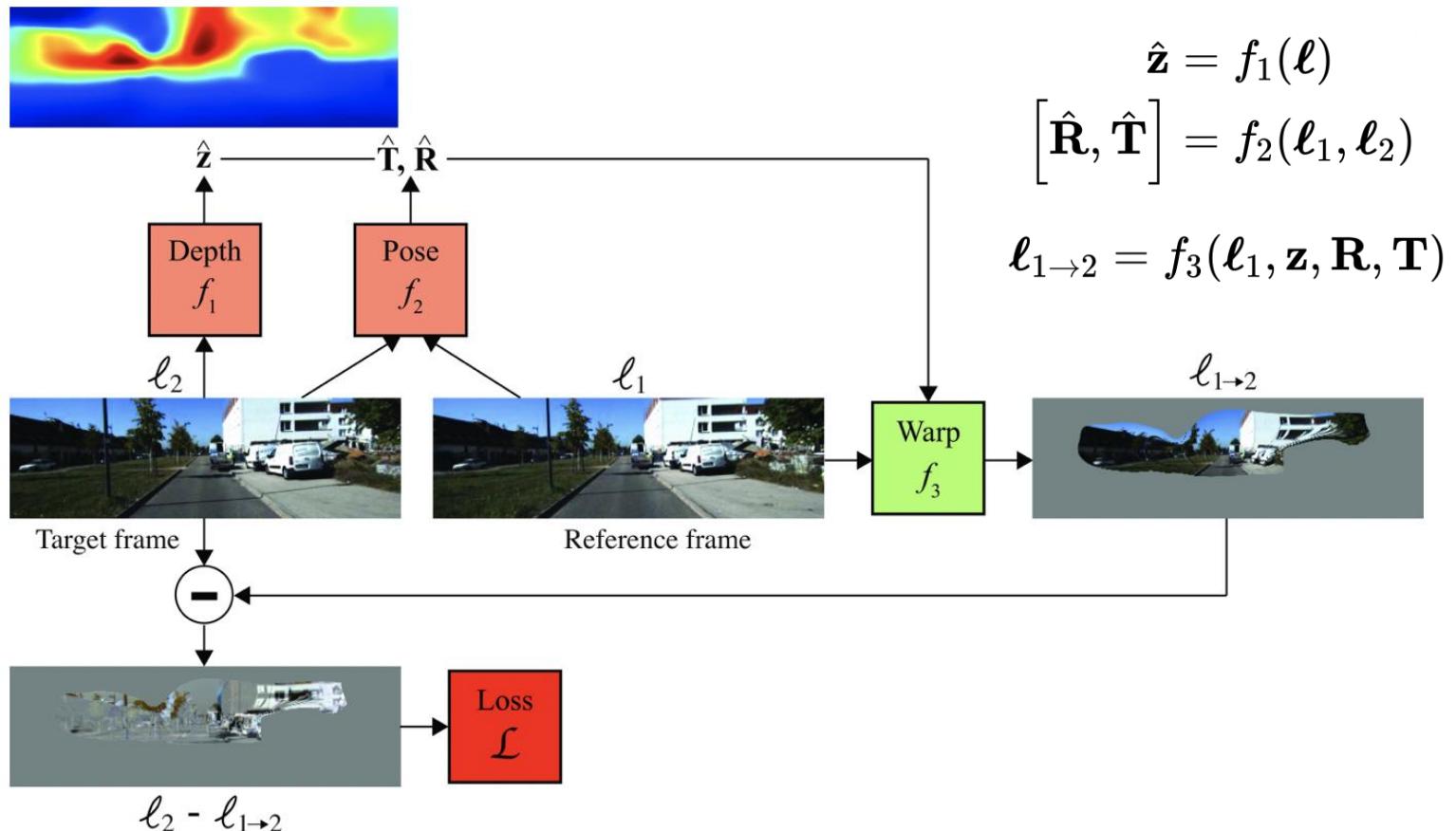
□ 有监督学习



- K. Xian, C. Shen, Z. Cao, H. Lu, Y. Xiao, R. Li, Z. Luo, Monocular relative depth perception with web stereo data supervision., in: Cvpr, 2018: pp. 311–320

基于学习的单目深度估计

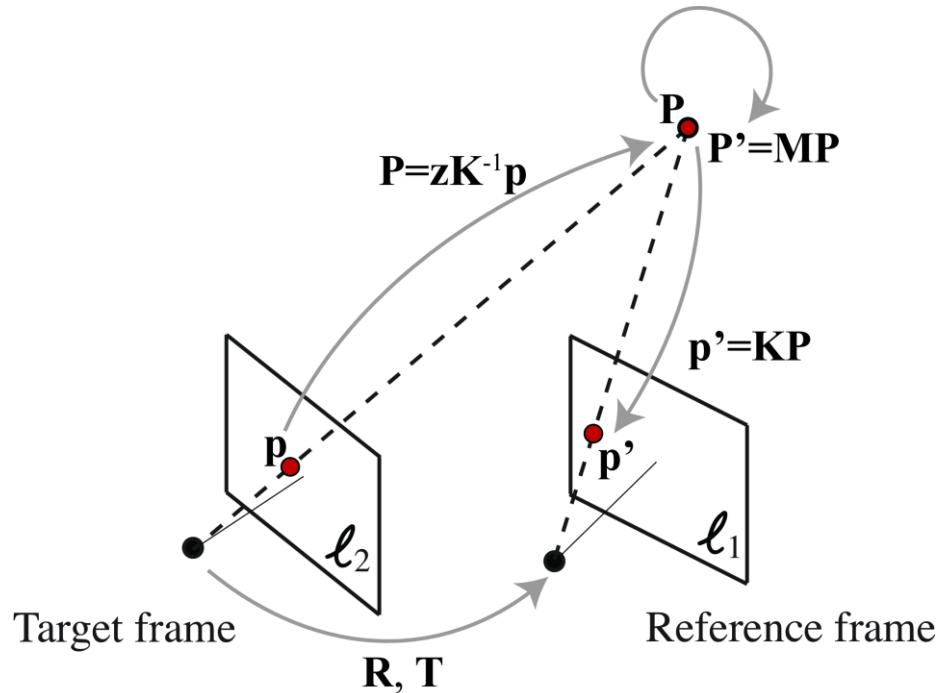
□ 自监督学习 (videos or multiple frames)



- T. Zhou, M. Brown, N. Snavely, D.G. Lowe, Unsupervised learning of depth and ego-motion from video. In CVPR, 2017.

基于学习的单目深度估计

□ 自监督学习



$$\ell_{1 \rightarrow 2}(p) = \ell_1(K\hat{M}\hat{z}(p)K^{-1}p)$$

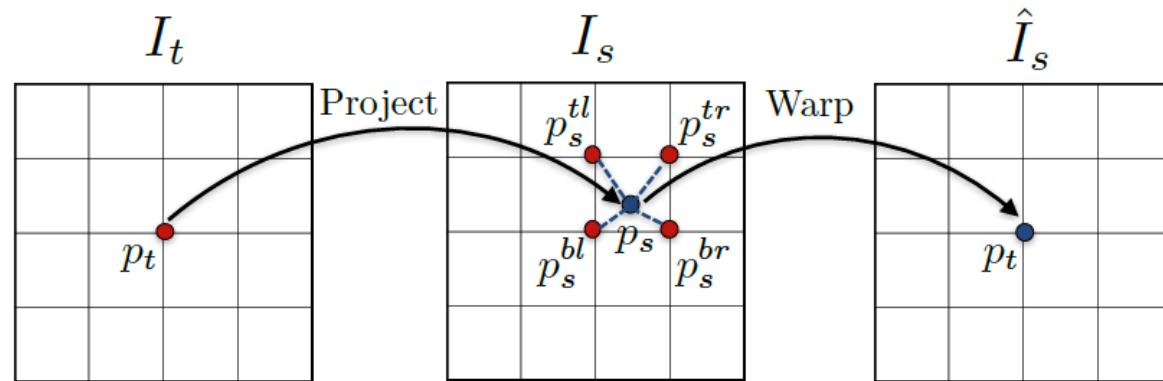
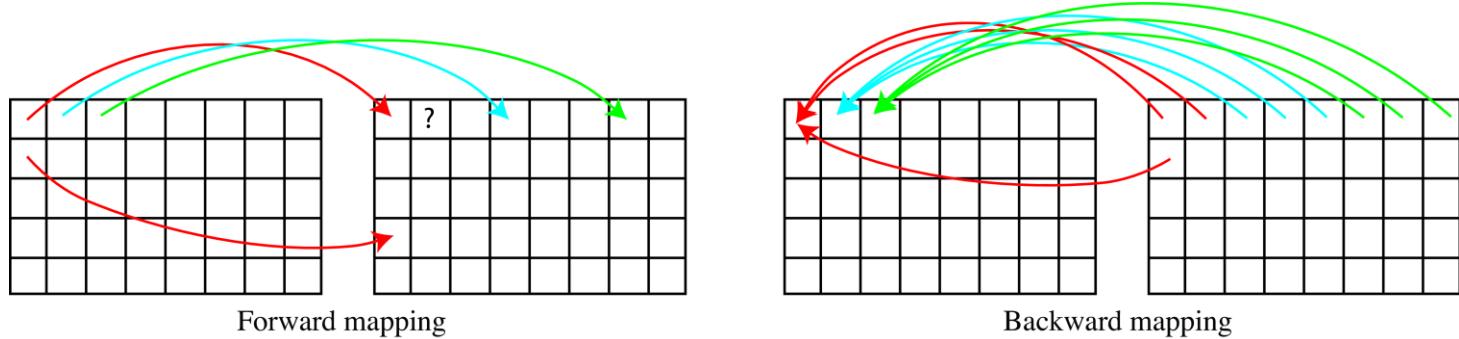
$$\mathcal{L} = \sum_{n,m} v(n,m) (\ell_{1 \rightarrow 2}[n,m] - \ell_2[n,m])^2$$

- T. Zhou, M. Brown, N. Snavely, D.G. Lowe, Unsupervised learning of depth and ego-motion from video. In CVPR, 2017.

基于学习的单目深度估计

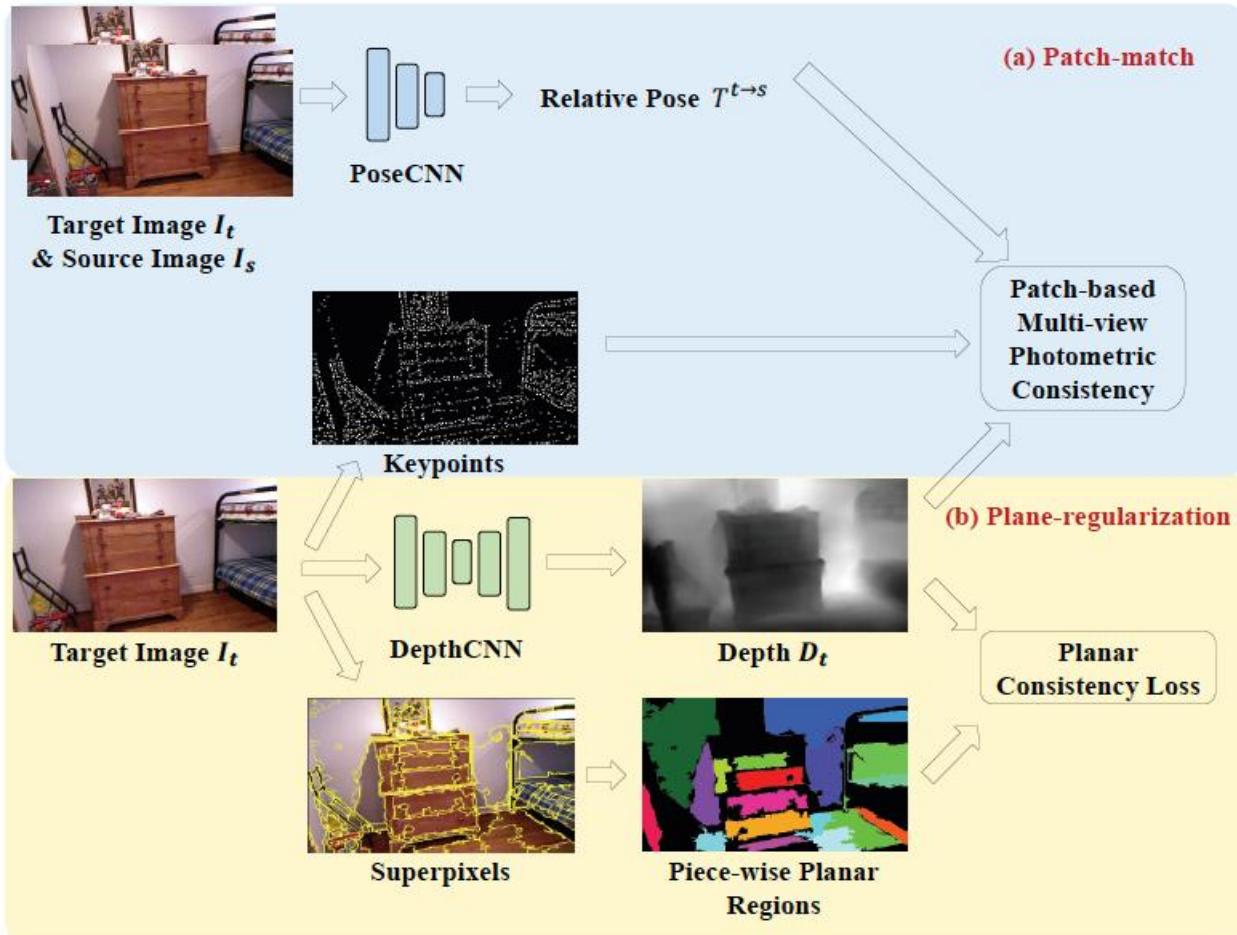
□ 自监督学习

■ 图像变形：前向 vs. 后向



基于学习的单目深度估计

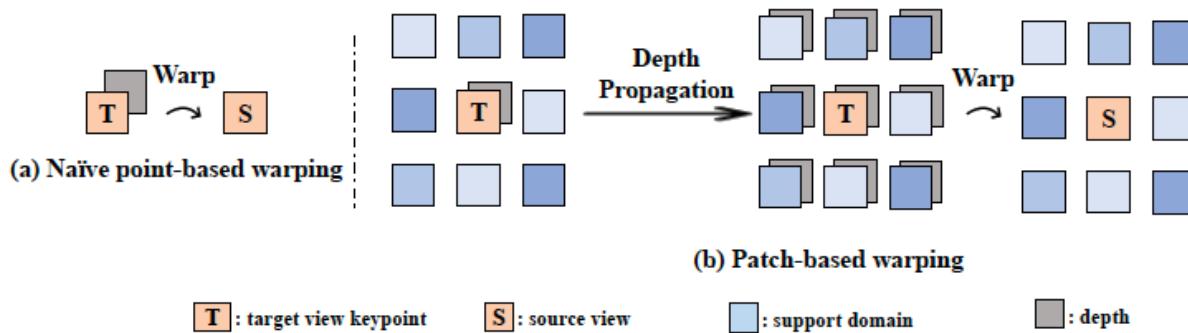
□ 自监督学习



基于学习的单目深度估计

□ 自监督学习

■ 基于区块的多视角光度测试一致性



■ 平面约束

$$p_n^{3D} = D(p_n)K^{-1}p_n, p_n \subseteq SPP_m$$

$$\text{平面方程} \quad A_m^\top p_n^{3D} = \mathbf{1}$$

$$L_{spp} = \sum_{m=1}^M \sum_{n=1}^N |D(p_n) - D'(p_n)| \quad \text{其中 } D'(p_n) = (A_m^\top K^{-1} p_n)^{-1}$$