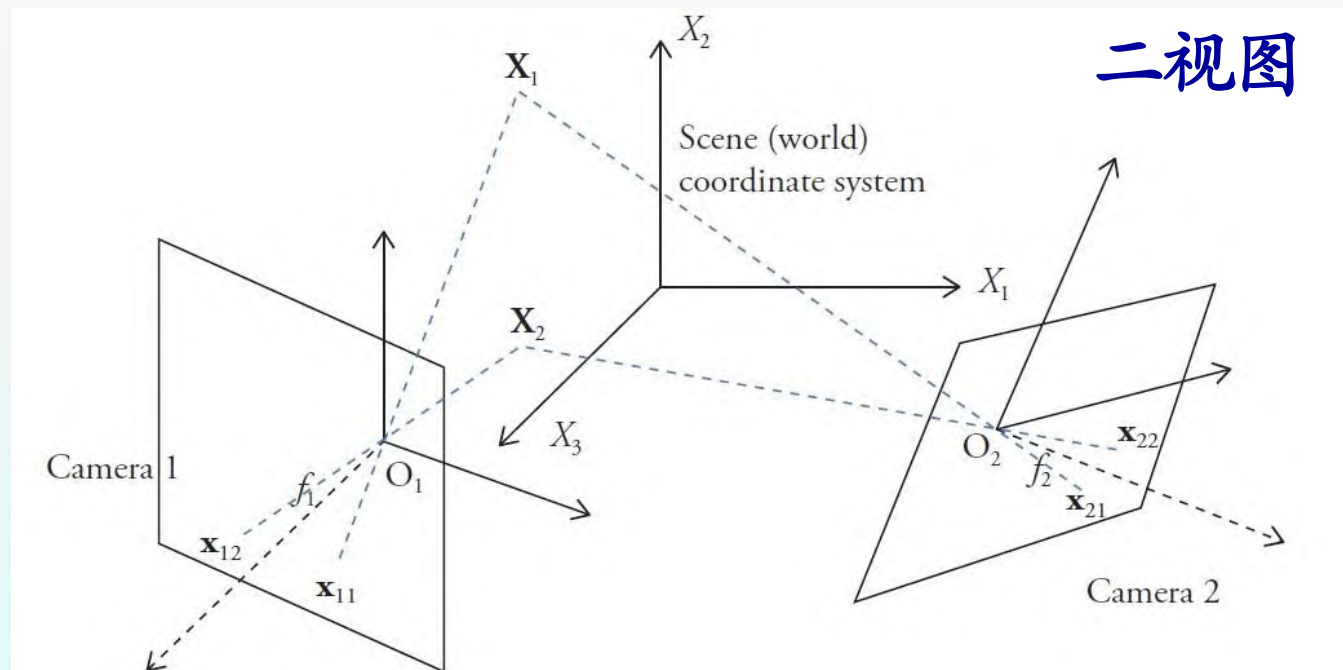


第14讲: 3D 运动/结构估计

- 3D 运动/结构估计
 - SFM 与 SFS
 - 双视角 vs. 多视角
 - 3D 重建分层法
- 相机标定
 - 一般的 3D 曲面 vs. 平面标定模式
- 运动仿射结构
 - 多帧因式分解方法
- 运动投影结构
 - 两帧
 - 多帧
- 欧几里得结构

SFM vs. SFS

- 运动恢复结构 (照相机运动, 静止场景)
- 立体恢复结构



运动的单目照相机:
立体视图:

Frame 1
左视图

Frame 2
右视图

运动恢复结构 (SFM)

- Koenderink和Van Doorn首先提出使用**两步求解**正交相机的SFM问题：**i)** 首先，通过至少两个未标定的视角，重建唯一的（可为任意仿射变换）3D场景表示，称作**仿射结构**；**ii)** 使用可用的度量（距离或角度）来唯一地确定**欧几里得结构**。
- Faugeras [Fau 95]将其扩展到**未标定相机的投影重建**，得到的投影是唯一的，最多有一次投影变换；并提出包括**投影，仿射和欧几里德**三个阶段的**分层3D重构法**，其中投影/仿射重建已经足够满足一些应用，而无需欧几里得重建所需要的费力的标定过程。

运动恢复结构 (cont'd)

- 当**标定不可用**时，对极几何由“**基本矩阵**”表示，也包括未知的相机标定。
- 在投影和仿射空间内分别定义**五个/四个选择点**，投影重建和仿射重建分别就是**求解这个稀疏点集相对于参考帧的3D坐标**。
- 在投影和仿射重建中，相对于欧几里德解，**未知的相机标定矩阵**（其中允许内部和外部相机参数在帧与帧之间变化）被融入到整个解的整体投影或仿射变形（模糊）中。

立体恢复结构 (SFS)

- **标定相机的立体（双视图）恢复结构**是重建稠密场景结构的最鲁棒的方法，因为对极几何对于所有像素都是成立的。
- **立体恢复结构（SFS）** 算法通常开始于相机标定和图像校正。
- **校正**是指将两个相机对准为共面，平行于连接两个相机中心的线。这是通过对两个图像应用单应性来实现的。
- 在图像校正之后，对应（**视差**）搜索被简化为在**水平线上的1D搜索**，因为所有的极线在校正的图像平面中是平行的。在第4.5节或第4.6节中讨论的运动估计方法可以用于1D视差估计。
- Scharstein和Szeliski [Sch 02]提出了超过20个立体匹配算法的比较评估。

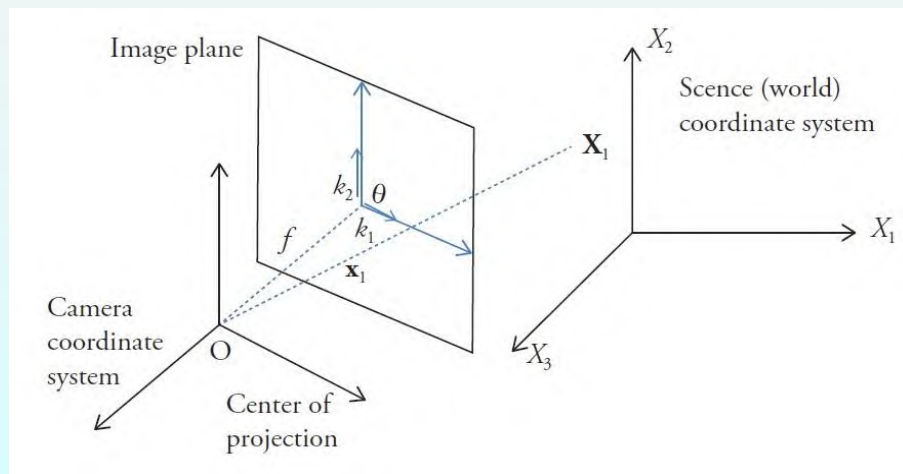
相机标定

- **相机标定**是指在投影相机模型中估计**内在**和**外在**相机参数（总共**11个自由参数**）：

$$\lambda \mathbf{x}_h = \begin{bmatrix} \lambda x_1 \\ \lambda x_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} f & k_1 & s & x_{1,0} \\ 0 & f & k_2 & x_{2,0} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ 1 \end{bmatrix} = \mathbf{K} [\mathbf{R} | \mathbf{t}] \mathbf{X}_h = \mathbf{P} \mathbf{X}_h$$

$$\mathbf{P} = \mathbf{K}[\mathbf{R} | \mathbf{t}].$$

- **内部相机矩阵K**具有**五个自由度**，其中 $(x_{1,0}, x_{2,0})$ 表示图像的中心， s 是偏斜参数， f 是相机的焦距， k_1 / k_2 表示像素宽高比。



- **外部相机参数**是具有三个自由度的旋转矩阵**R**和具有三个参数的平移矢量**t**，分别对相机坐标系相对于场景（世界）坐标系的旋转和平移进行建模。

相机标定 (cont'd)

- **相机标定技术**可以被分类为使用已知参考对象的**预标定**和依赖于来自实际视频场景而没有任何参考对象的点对应的**自动标定**（或**自标定**）。
- **预标定方法**可以利用来自不同视点的**已知3D参考对象**[Tsa 87]或**平面目标**[Zha 00]的多个图像。
- 预标定技术可以分为参数的直接估计与**两步估计**：
 - 首先估计相机投影矩阵 $\mathbf{P} = \mathbf{K} [\mathbf{R} \mid \mathbf{t}]$
 - 然后从投影矩阵恢复内外参数。
- 应尽可能**优先选择预标定**，因为自标定不能总是达到与预标定相同的精度水平。
- 在一些情况下，仅可能执行**部分预标定**以确定**相机中心**，**像素纵横比**和**偏斜度**参数，因为焦距和相机旋转和平移在摄录（捕获）期间会发生变化。

相机标定 (cont'd)

- 使用3D参考对象常用的**预标定方法**包括以下步骤：
 - 1) **检测**对应于已知世界坐标 \mathbf{X}_i 的3D参考对象点的**图像特征点 \mathbf{x}_i** ;
 - 2) 利用至少**6个世界图像点对应**，采用线性最小二乘法求解 $\mathbf{x} = \mathbf{P}\mathbf{X}$ ，估计相机投影矩阵 \mathbf{P} ;
 - 3) 根据矩阵 \mathbf{P} 各元素的闭合函数形式**计算内外相机参数 \mathbf{K} ， \mathbf{R} 和 \mathbf{t}** [Zha 04];
 - 4) 通过**非线性优化改进 \mathbf{K} ， \mathbf{R} 和 \mathbf{t}**

$$\min_{\mathbf{K}, \mathbf{R}, \mathbf{t}} \sum_i \|\mathbf{x}_i - \mathbf{P}\mathbf{X}_i\|^2$$

从步骤3中的估计开始。

- 或者，我们可以首先通过非线性优化（即在步骤2之后执行步骤4）来改进 \mathbf{P} ，然后从更新的 \mathbf{P} 确定 \mathbf{K} ， \mathbf{R} 和 \mathbf{t} 。

仿射重建

- **仿射结构重建问题** [Koe 90]: 给定由未标定仿射相机捕获的两个视图中**至少五个点对应**:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{f}{Z} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \mathbf{M} \mathbf{X}$$

- 任意地选择其中**四个不在退化配置上的点**，以定义**仿射坐标系**
- 在这四个点定义的仿射坐标系中**重建第五点**（或已知对应的其他任意点）。
- 与欧几里德结构不同，仿射结构通过未知的仿射变换，**改变了度量距离和角度**，但**保留了平行结构**。
- 使用相机标定参数或部分场景度量测量，可以在分层法的第二步中计算出该变换，以获得可达**某个尺度因子的欧几里德结构**。
- Koenderink-Van Doorn [Koe 90]为**双视角仿射结构重建问题**提供了几何解。

多视角仿射重建

- **因子分解方法** [Tom 92]: 当有两个以上的视角时, 一次性估计所有相机矩阵和特征的仿射结构。

假设在由 \mathbf{x}_{ij} , $i = 1, \dots, M$, $j = 1, \dots, N$ 表示的 M 个视图中的 N 个特征点之间具有对应信息。在仿射相机和刚性运动的情况下, 包含了连续视图中所有对应点的图像坐标的 $2M \times N$ 测量矩阵 \mathbf{W} , 其秩为3。这是由于图像平面中的特征点的位置将会受到刚性3D运动的约束, 因此 \mathbf{W} 发生降秩。

- **矩阵 \mathbf{W}** 可以表示为相机运动矩阵 \mathbf{M} 和仿射结构矩阵 \mathbf{X} 的乘积:

$$\mathbf{W} = \begin{bmatrix} \mathbf{x}_{11} & \mathbf{x}_{12} & \cdots & \mathbf{x}_{1N} \\ \mathbf{x}_{21} & \mathbf{x}_{22} & \cdots & \mathbf{x}_{2N} \\ \vdots & & \ddots & \vdots \\ \mathbf{x}_{M1} & \mathbf{x}_{M2} & \cdots & \mathbf{x}_{MN} \end{bmatrix}_{2M \times N} = \begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \\ \vdots \\ \mathbf{M}_M \end{bmatrix}_{2M \times 3} [\mathbf{X}_1 \quad \mathbf{X}_2 \quad \cdots \quad \mathbf{X}_N]_{3 \times N} = \mathbf{M} \mathbf{X}$$

其中每个相机矩阵 \mathbf{M}_i 为 2×3 , \mathbf{x}_{ij} 为 2×1 , \mathbf{X}_i 为 3×1 .

多视角仿射重建 (cont'd)

- Tomasi和Kanade [Tom 92]提出使用**奇异值分解 (SVD)** 对测量矩阵 **W** 进行**因式分解**,

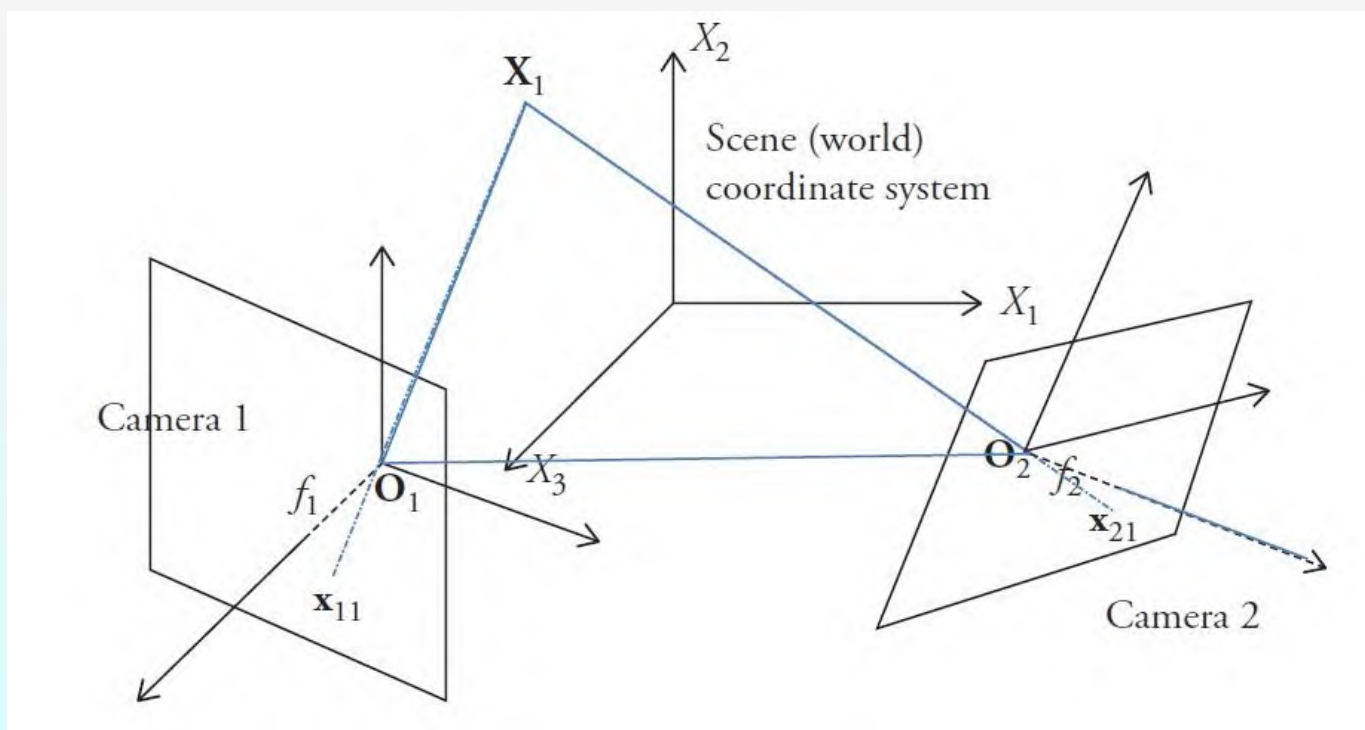
$$W = UDV^T$$

其中 **$M = U$** 对应于运动 (相机) 矩阵, 并且 **$X = DV^T$** 对应于通过仿射结构向量相对于以这些点中的一个为中心的坐标系形成的矩阵。

- 多视点仿射重建方法可概括为:
 - 1) 构建**测量矩阵 **W**** ,
 - 2) 执行 **W 的SVD分解**以恢复 **M** 和 **X** .
 - 3) **恢复未知的仿射变换**, 使用相机标定参数或一些关于场景的度量测量来获得达到某个比例因子的欧几里得结构。

投影重建

- **对极几何**：连接 \mathbf{x}_{1j} 的摄像机中心 \mathbf{O}_1 和 \mathbf{x}_{2j} 及其摄像机中心 \mathbf{O}_2 的线在点 \mathbf{X}_i 相交，或者，从 \mathbf{O}_1 到 \mathbf{x}_{1j} 的线，从 \mathbf{O}_2 到 \mathbf{x}_{2j} 的线以及从 \mathbf{O}_1 到 \mathbf{O}_2 的线都是共面的。



投影重建 (cont'd)

- **基本矩阵F**在两个视图之间捕获此**对极几何结构**，它们仅取决于**相机参数和姿态**，在齐次坐标中的代数表达式中，有：

$$\mathbf{x}_{1j}^T \mathbf{F} \mathbf{x}_{2j} = 0, \quad j = 1, \dots, N$$

可以看出，基本矩阵可以用**内在**和**外在的相机参数**表示为 $\mathbf{F} = (\mathbf{K}_1^{-1})^T \mathbf{E} \mathbf{K}_2^{-1}$ ，其中 $\mathbf{E} = \mathbf{t}_x \mathbf{R}$ ，而

$$\mathbf{t}_x = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix}$$

- 当我们无法访问内部标定参数 \mathbf{K}_1 和 \mathbf{K}_2 时，可以从 \mathbf{F} 获得的重建将达到实际欧几里德结构的未知投影变换 \mathbf{H}^{-1} 。

投影重建 (cont'd)

三维投影结构二维重建的完整算法可概括为：

- **像素对应**：在两个视图之间找到至少 $N = 8$ 个像素对应。
- **归一化**[Har 97]：归一化像素坐标 $\tilde{\mathbf{x}}_{ij} = \mathbf{T}_i \mathbf{x}_{ij}$ ， $i = 1, 2$ ， $j = 1, \dots, N$ ，其中仿射映射 \mathbf{T}_i 的计算如下：
 - 1) 将坐标的原点**移动到平均值** $\bar{\mathbf{x}}_i = \sum_{j=1}^N \mathbf{x}_{ij}$ ， $i = 1, 2$ 。
 - 2) **缩放平移后像素**， $\tilde{\mathbf{x}}_{ij} = k(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)$ ，使得它们与原点的均方根距离为 $\sqrt{2}$ 。

投影重建 (cont'd)

- **基本矩阵**: 给定归一化对应点, 估计基本矩阵 $\tilde{\mathbf{F}}$.

1) 令 $\tilde{\mathbf{x}}_{ij} = [u_{ij} \ v_{ij} \ w_{ij}]$, $i = 1, 2, j = 1, \dots, N$; 构造方程: $\mathbf{A}\mathbf{f} = \mathbf{0}$,

其中 $\mathbf{f} = [\tilde{F}_{11} \ \tilde{F}_{12} \ \dots \ \tilde{F}_{33}]^T$ 为包含 3×3 基本矩阵 $\tilde{\mathbf{F}}$ 元素的 9×1 向量,
 $N \times 9$ 矩阵 \mathbf{A} 的第 j 行具有以下形式:

$$[u_{2j}u_{1j} \ u_{2j}v_{1j} \ u_{2j}w_{1j} \ v_{2j}u_{1j} \ v_{2j}v_{1j} \ v_{2j}w_{1j} \ w_{2j}u_{1j} \ w_{2j}v_{1j} \ w_{2j}w_{1j}]$$

注意到存在唯一解的条件是矩阵 \mathbf{A} 的**秩必须为8**。

2) 通过求解下式可得到该方程组的**非平凡解**:

$$\begin{aligned} \min & \|\mathbf{A}\mathbf{f}\| \\ \text{subject to } & \|\mathbf{f}\| = 1 \end{aligned}$$

解是 $\mathbf{A}^T\mathbf{A}$ 对应于最小特征值的**特征向量**。

3) 归一化前的基本矩阵 \mathbf{F} 由 $\mathbf{F} = \mathbf{T}_2^T \tilde{\mathbf{F}} \mathbf{T}_1$ 给出。

投影重建 (cont'd)

- **相机投影矩阵**：给定 F ，恢复相机矩阵 $P_1 = [I \ 0]$ 和 $P_2 = [R \ t]$ 。

1) 计算**奇异值分解** $F = UDV^T$ 。由于 F 的秩为2，所以在没有噪声的情况下，它应该具有两个非零奇异值 $D \approx \text{diag}(a, b, 0)$ 。

2) 观察[Har 04]: $F = (UZU^T)(UY^T\hat{D}V^T) = S M,$

$$\text{其中: } Y = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad Z = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad D \approx \text{diag}(a, b, c).$$

c 的值可以任意设定，常用的选择是 $c = (a + b) / 2$ 。

3) 则相机矩阵是 $P_1 = [I \ 0]$ 和 $P_2 = [M \ u_3]$ ，其中 u_3 是 U 的第三列。注意： $u_3^T F = 0$ ，即 u_3 是 F 的左零空间的生成向量。

投影重建 (cont'd)

- **三角测量**: 由于估计不准确, 导致连接 \mathbf{x}_{1j} 和 \mathbf{x}_{2j} 及其各自相机中心的线在3D空间中不相交, 因此我们通过求解下式来估计3D点 $\hat{\mathbf{X}}_j$:
$$\text{Min}_{\mathbf{x}_j} \left\| \mathbf{x}_{1j} - \mathbf{P}_1 \hat{\mathbf{X}}_j \right\|^2 + \left\| \mathbf{x}_{2j} - \mathbf{P}_2 \hat{\mathbf{X}}_j \right\|^2$$

这对噪声是鲁棒的[Har 04]。此步骤需要一个迭代优化的过程。

多视角投影因式分解

- Sturm-Triggs [Stu 96]等[Ole 07]将基于仿射重建的扩展因式分解法用于投影重建。**M个视图**和**N个特征点**的投影方程以矢量矩阵形式表示为：

$$\mathbf{O} = \begin{bmatrix} \lambda_{11}\mathbf{x}_{11} & \lambda_{12}\mathbf{x}_{12} & \cdots & \lambda_{1N}\mathbf{x}_{1N} \\ \lambda_{21}\mathbf{x}_{21} & \lambda_{22}\mathbf{x}_{22} & \cdots & \lambda_{2N}\mathbf{x}_{2N} \\ \vdots & & \ddots & \vdots \\ \lambda_{M1}\mathbf{x}_{M1} & \lambda_{M2}\mathbf{x}_{M2} & \cdots & \lambda_{MN}\mathbf{x}_{MN} \end{bmatrix}_{3M \times N} = \begin{bmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \\ \vdots \\ \mathbf{P}_M \end{bmatrix}_{3M \times 4} \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_N \end{bmatrix}_{4 \times N}$$

- 其中比例因子 λ_{ij} 称为**投影深度**。请注意，每个视图都要求每个点都可见（无遮挡），这样观察矩阵 \mathbf{O} 就不会缺少元素。我们可以将矩阵 \mathbf{O} 表示为矩阵的Hadamard（元素）乘积 \odot ：

$$\mathbf{O} = \mathbf{\Lambda} \odot \mathbf{W}, \text{ 其中 } \mathbf{\Lambda} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1N} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2N} \\ \vdots & & \ddots & \vdots \\ \lambda_{M1} & \lambda_{M2} & \cdots & \lambda_{MN} \end{bmatrix} \quad \mathbf{W} = \begin{bmatrix} \mathbf{x}_{11} & \mathbf{x}_{12} & \cdots & \mathbf{x}_{1N} \\ \mathbf{x}_{21} & \mathbf{x}_{22} & \cdots & \mathbf{x}_{2N} \\ \vdots & & \ddots & \vdots \\ \mathbf{x}_{M1} & \mathbf{x}_{M2} & \cdots & \mathbf{x}_{MN} \end{bmatrix}$$

多视投影分解 (cont'd)

- **投影因式分解**旨在同时解决未知投影深度 Λ ，未知摄像机 \mathbf{P} 和未知结构 \mathbf{X} 。
- 理想情况下，**矩阵 \mathbf{O} 必须具有秩4**。但是，由于投影深度 λ_{ij} 是未知的，所以如果它们被任意设置，则该秩为4的属性不会被满足；因此，我们需要准确的初始化投影深度。
- Sturm-Triggs [Stu 96]提出了一种**迭代算法**，通过**交替求解问题**：（1）给定 Λ ，通过SVD因式分解求解 \mathbf{P} 和 \mathbf{X} ；（2）给定 \mathbf{P} 和 \mathbf{X} ，通过最小二乘拟合求解 Λ 。在这两个步骤之间交替，直到收敛。在他们的原始论文中，Sturm-Triggs从连续的两视图重建中对投影深度进行初始化。
- 已经注意到，如果真正的解与**接近于其仿射近似**[Har 04]，具有上述初始化的Sturm-Triggs方法可较好工作。

多视投影分解 (cont'd)

- 原始方法的**扩展方法**可分为以下步骤:
- 对于奇次观察矩阵 \mathbf{W} , 初始化 $\Lambda^{(0)}$, 使得所有 $\lambda_{ij}^{(0)} = 1$.
- 循环操作, $k = 0, \dots, N$
 - (a) 通过奇异值分解计算最接近的秩为4的矩阵 $\mathbf{O}^{(k)}$ 到 $\Lambda^{(k)} \odot \mathbf{W}$ 。
令 $\Lambda^{(k)} \odot \mathbf{W} = \mathbf{U} \mathbf{D} \mathbf{V}^T$ 。将 \mathbf{D} 定义为通过保留 \mathbf{D} 的前四个 (最大) 对角项而获得的对角矩阵, 并将其余部分设置为零。那么,
 $\mathbf{O}^{(k)} = \mathbf{U} \mathbf{D} \mathbf{V}^T$
 - (b) 如果 $k = N$, 则转到步骤3。否则, 计算权重 $\lambda_{ij}^{(k+1)}$ 的新矩阵 $\Lambda^{(k+1)}$, 使得 $\Lambda^{(k+1)} \odot \mathbf{W}$ 在Frobenius范数下尽可能接近 $\mathbf{O}^{(k)}$ 。
- 计算**因式分解** $\mathbf{O}^{(N)} = \mathbf{P} \mathbf{X}$, 以获得可分别提供相机矩阵和点位置的 \mathbf{P} 和 \mathbf{X} 。

聚束调整

- 由于双视角和多视角问题中特征点对应 \mathbf{x}_{ij} 的**准确性有限**，因而无法准确地满足投影方程 $\lambda_{ij}\mathbf{x}_{ij} = \mathbf{P}_i\mathbf{X}_j$ 。因此，通过使用简单线性估计的投影矩阵 \mathbf{P}_i 和3D点 \mathbf{X}_j 可能包含各种错误。
- 假设 \mathbf{P}_i 和 \mathbf{X}_j 的测量误差是高斯的且彼此独立，最大似然（ML）估计需要求解以下**非线性最小二乘问题**：

$$\min_{\mathbf{P}_i, \mathbf{X}_j} \sum_i \sum_{j=1}^N \|\mathbf{x}_{ij} - \mathbf{P}_i \mathbf{X}_j\|^2$$

这被称为“**聚束调整**”。用于最小化该代价函数的常用迭代算法是Levenberg-Marquardt方法[Tri 99, Har 04]。

- 为了融合到全局最优解，它需要一个**很好的初始解**，可用双视角或多视角解来找到。

欧氏结构重建

- **欧几里德重建**的经典方法需要知道**相机标定矩阵**。
在双视图情况下，给定标定矩阵 \mathbf{K}_1 和 \mathbf{K}_2 ，可以首先计算：

$$\check{\mathbf{x}}_{ij} = \mathbf{K}_i^{-1} \mathbf{x}_{ij}, i = 1, 2$$

- 则**极线约束**可写成：

$$\mathbf{x}_{1j}^T (\mathbf{K}_1^{-1})^T \mathbf{E} \mathbf{K}_2^{-1} \mathbf{x}_{2j} = 0,$$

可以被表达为：

$$\check{\mathbf{x}}_{1j}^T \mathbf{E} \check{\mathbf{x}}_{2j} = 0, \quad j = 1, \dots, N$$

- 因此，可以**直接**从缩放的像素对应**恢复本质矩阵E**和3D欧几里得重建（达到全局尺度）。

欧氏结构重建 (cont'd)

- 或者，在分层方法中，首先从未标定的像素对应度**计算投影重建**，然后通过考虑标定真实值或相机标定信息将其**升级为仿射或欧氏距离**。这可以使用不同方式实现：
 - **相机标定**：给定相机矩阵，投影模糊度可以解析为全局比例因子。
 - **部分相机标定**：聚束调整期间可能会执行约束，例如已知焦距，或相机具有相同的内部参数。
 - **自动标定**：自动标定的自动方法通常首先计算仿射重建，然后更新为欧几里德重建并完全确定摄像机标定参数[Har 04]。
 - **标定真实值3D坐标**：需要知道至少5个真实控制点的3D欧氏坐标[Har 04]。

习题

- 4.2 (假定焦距 $f=1$)
- 4.4
- 4.6
- 4.9 (注意本题是如何判定遮挡区域, 并非讨论遮挡区域是如何产生的)
- 大作业: Matlab Exercises 中任选一道题。